



# Integration of Heterogeneous Data

Gary Bader

Terrence Donnelly Center for Cellular and Biomolecular Research  
(CCBR)

BBDMR, SLRI

University of Toronto

MoGen Topic Course - Mar.31.2008

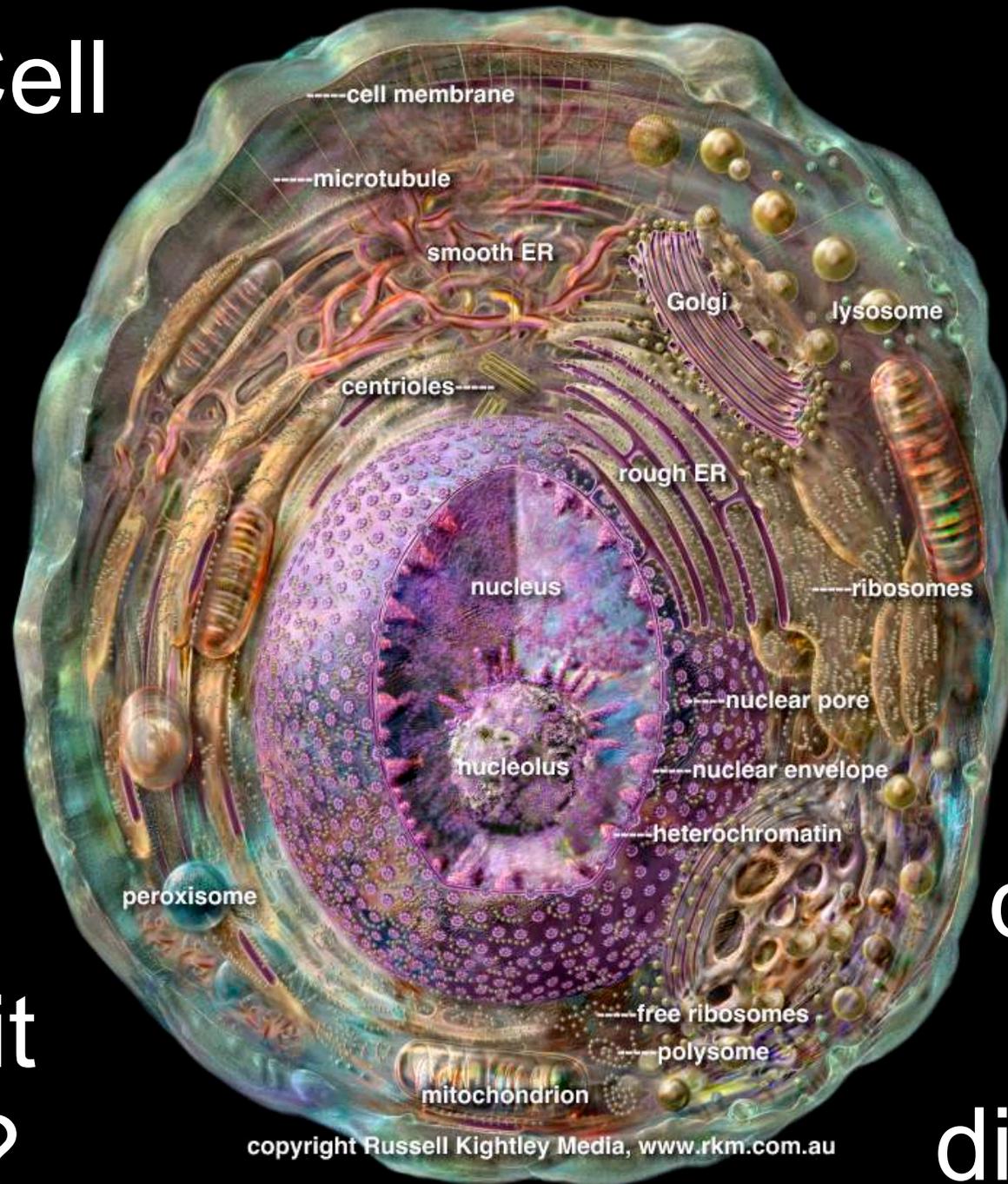
<http://baderlab.org>



# Outline

- Networks as scaffold for data integration
- Statistical integration
  - Gene function prediction
- Database and semantic integration
  - Biological pathways
- Visualizing integrated data with Cytoscape
- Example: Breast cancer classification

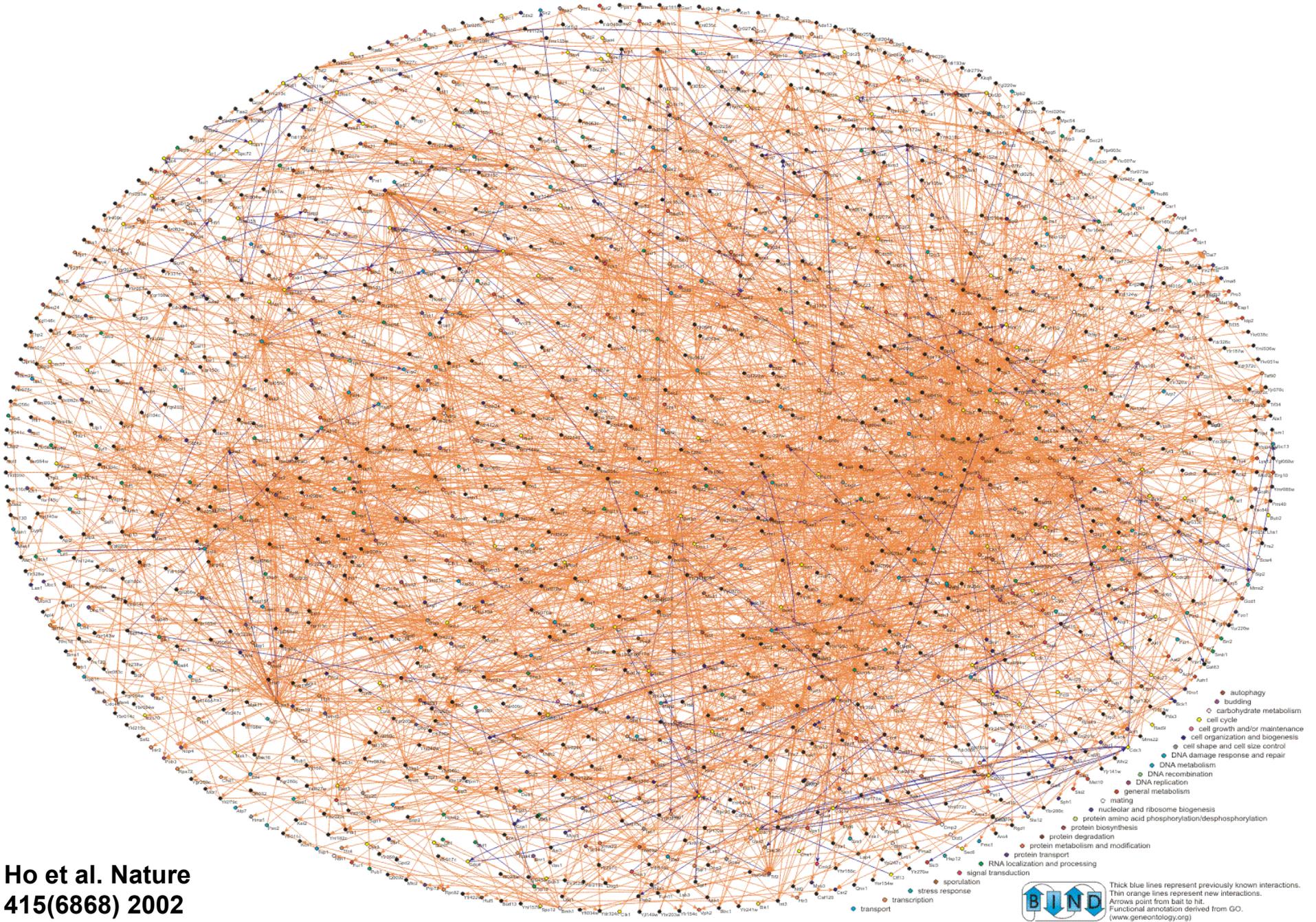
# The Cell



How  
does it  
work?

How  
does it  
fail in  
disease?

# Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry



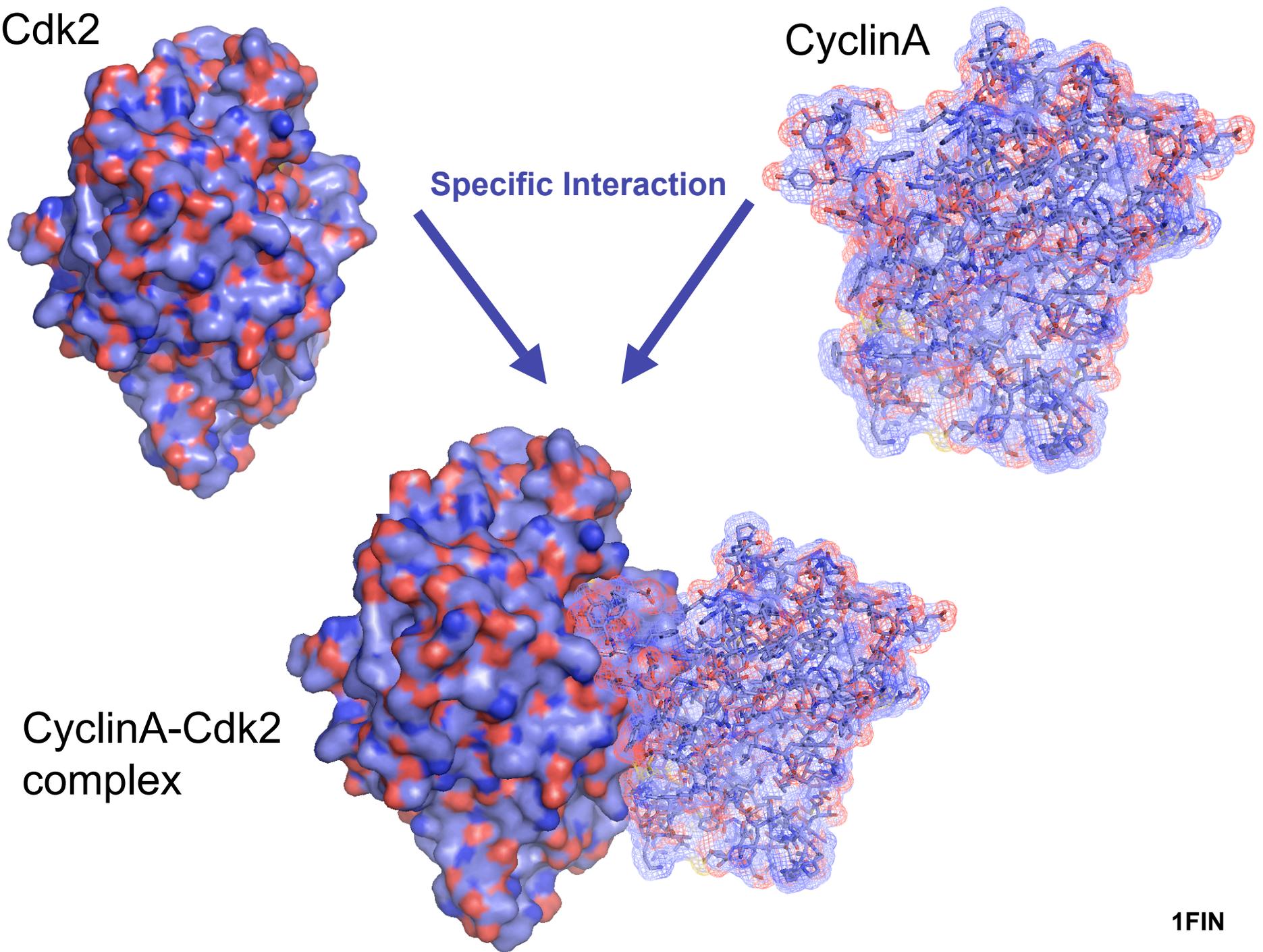
Ho et al. Nature  
415(6868) 2002

Cdk2

CyclinA

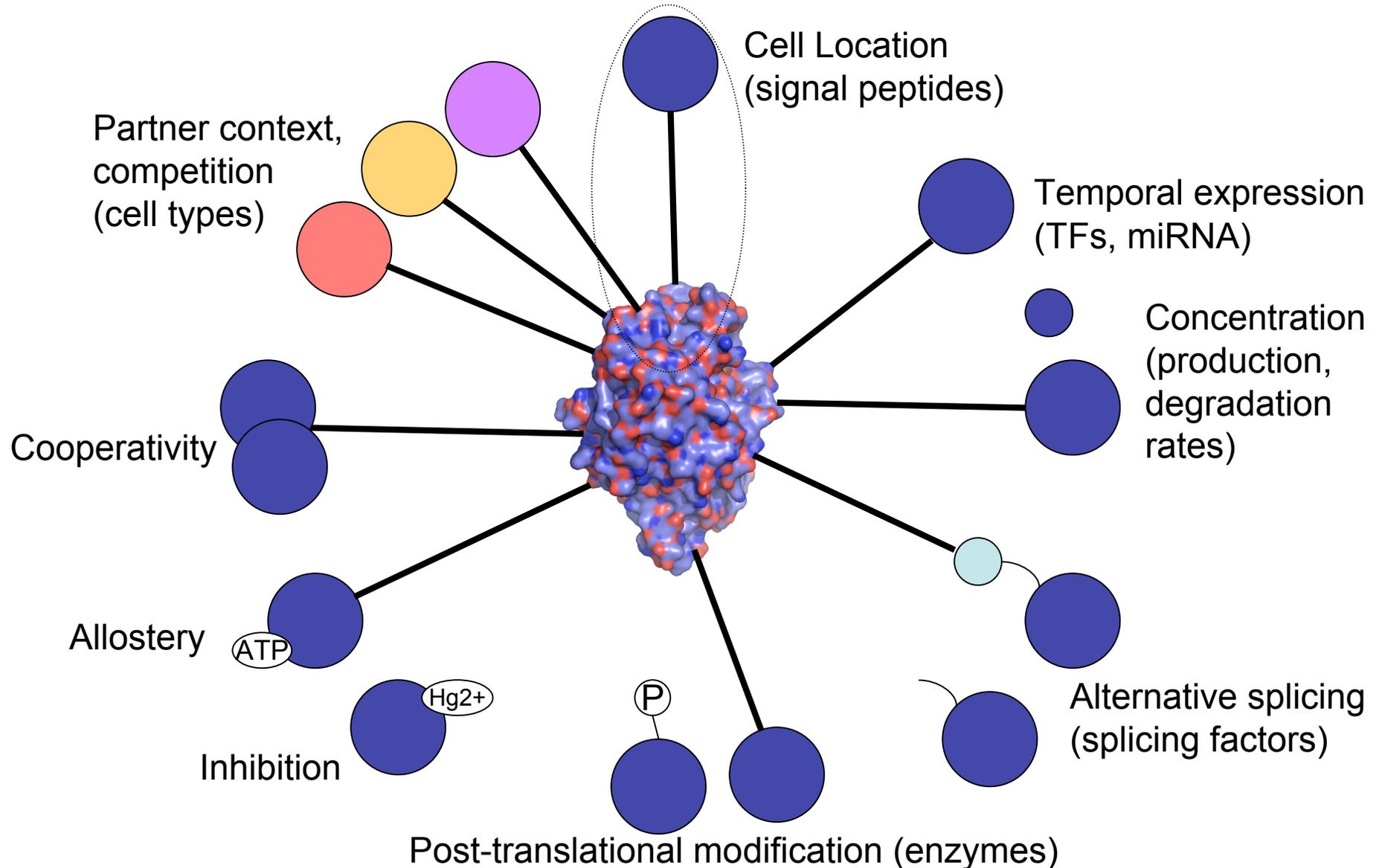
Specific Interaction

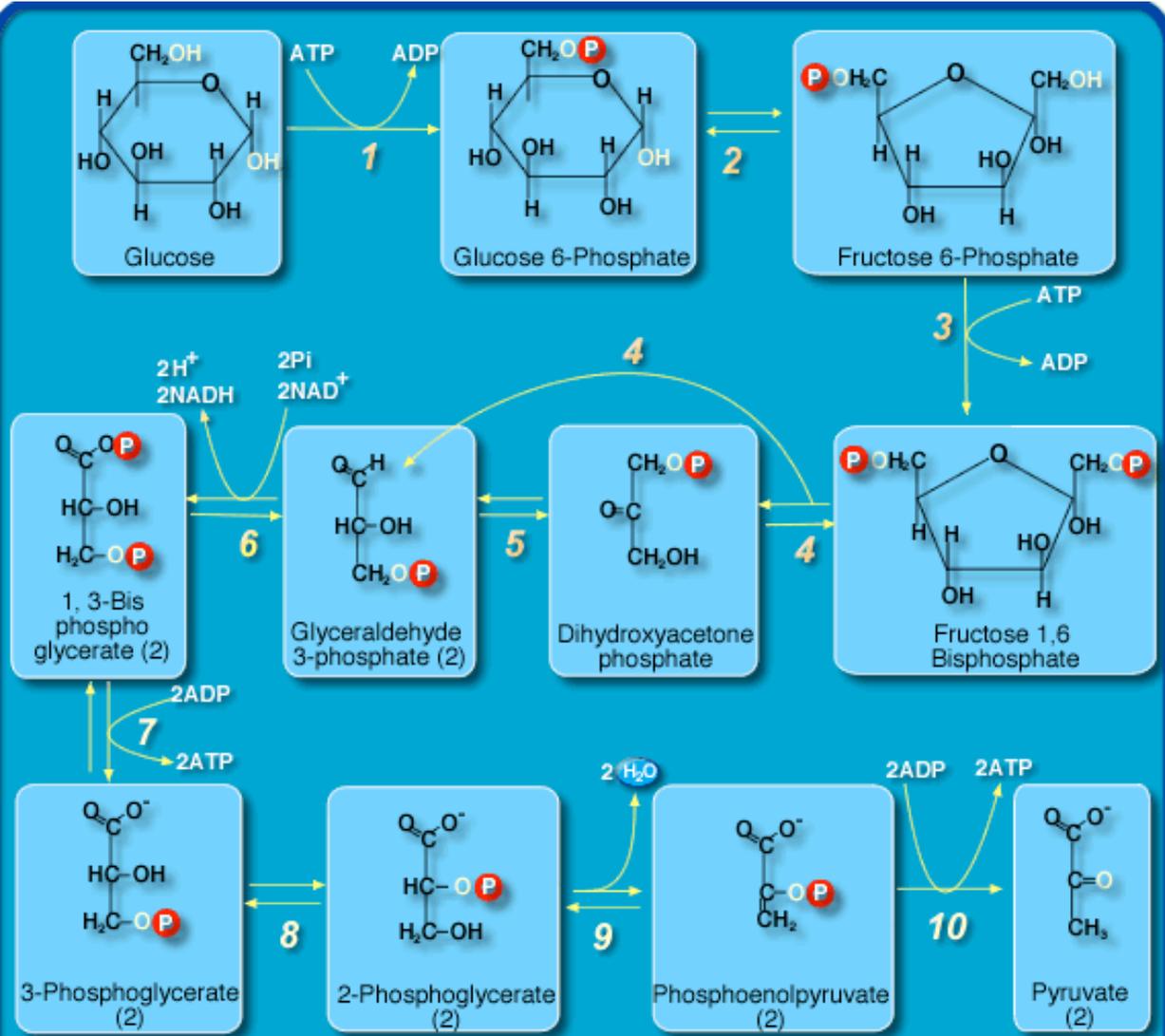
CyclinA-Cdk2  
complex



1FIN

# Towards More Biologically Relevant Networks





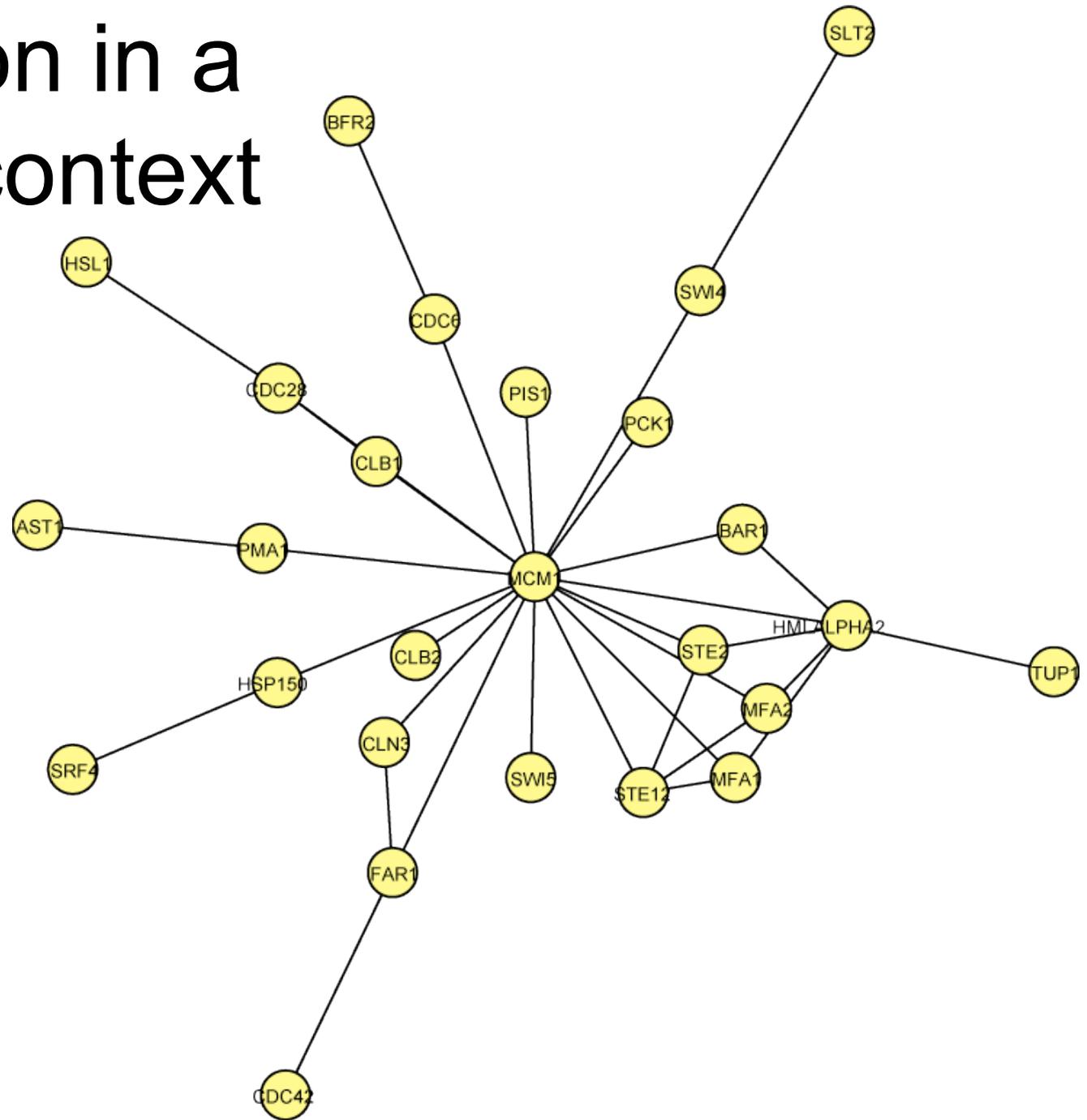
**ENZYMES**

- |   |  |  |
|---|--|--|
| <ul style="list-style-type: none"> <li>1 Hexokinase</li> <li>2 Glucose Phosphate Isomerase</li> <li>3 Phosphofructokinase</li> <li>4 Fructose diphosphate aldolase</li> </ul> | <ul style="list-style-type: none"> <li>5 Triose phosphate Isomerase</li> <li>6 Glyceraldehyde Phosphate Dehydrogenase</li> </ul> | <ul style="list-style-type: none"> <li>7 Phosphoglycerate Kinase</li> <li>8 Phosphoglyceromutase</li> <li>9 Enolase</li> <li>10 Pyruvate Kinase</li> </ul> |
|---|--|--|

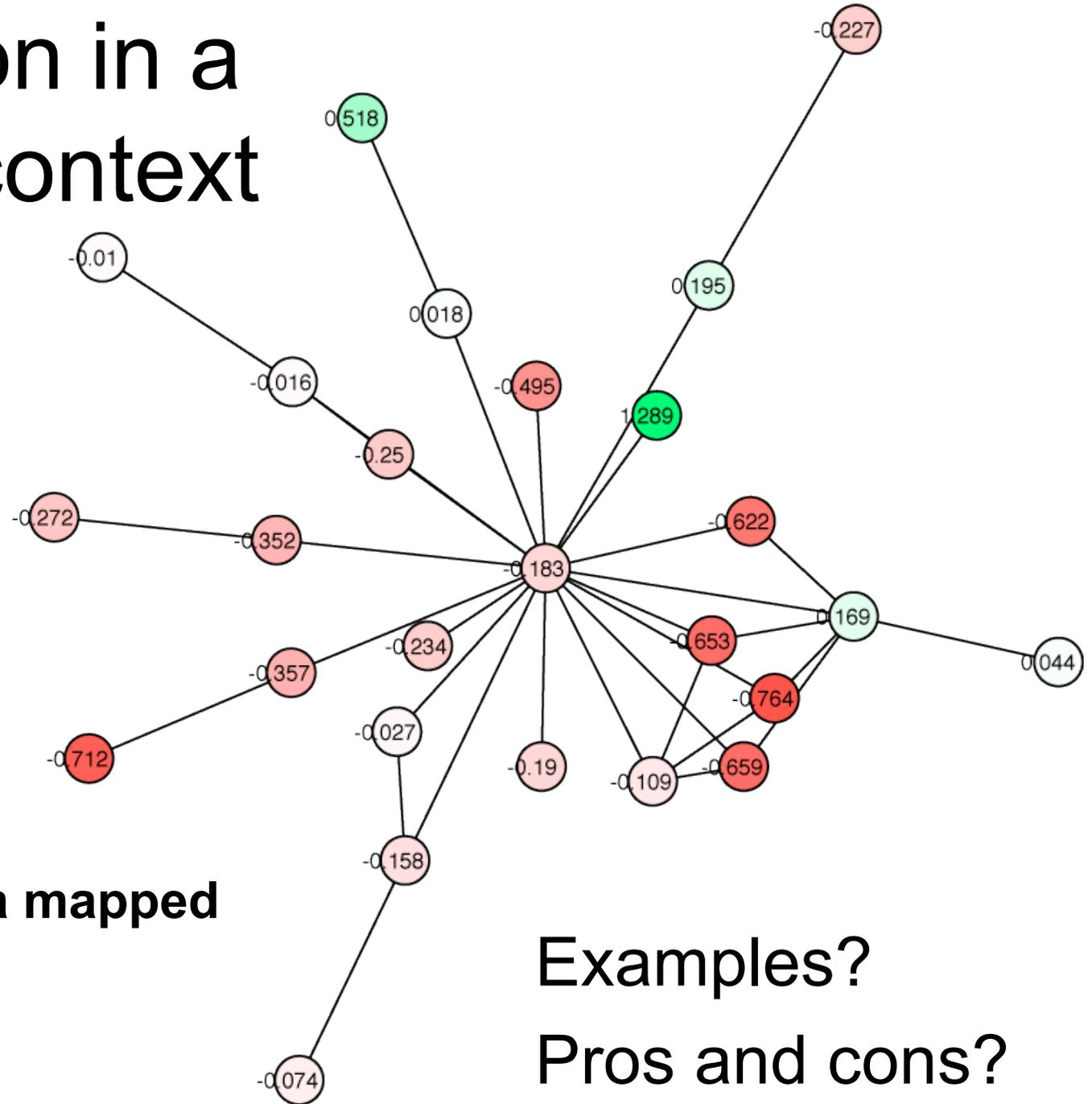




# Integration in a network context



# Integration in a network context



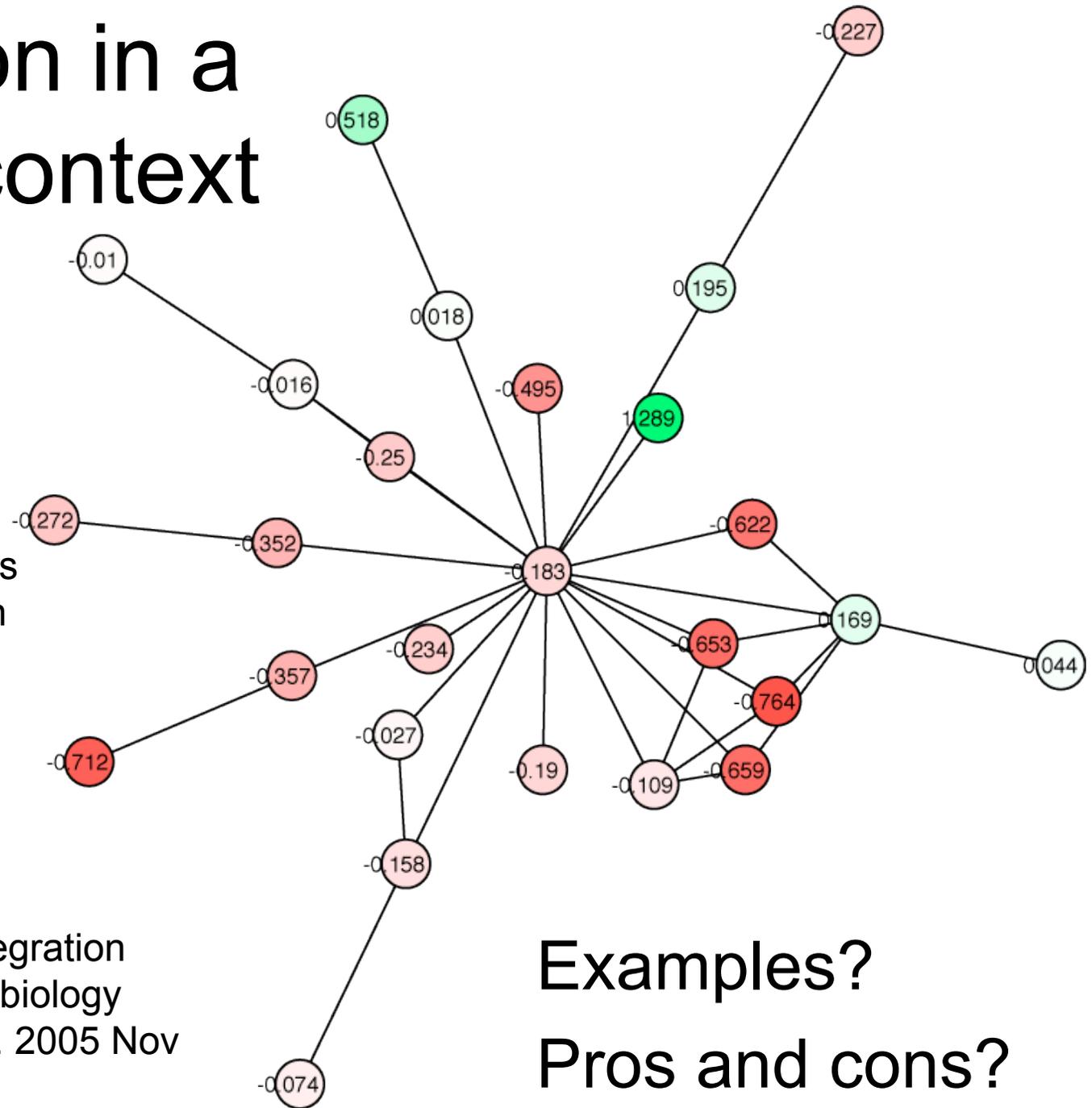
# Integration in a network context

## Advantages:

- Broader coverage
- Error reduction

## Challenges:

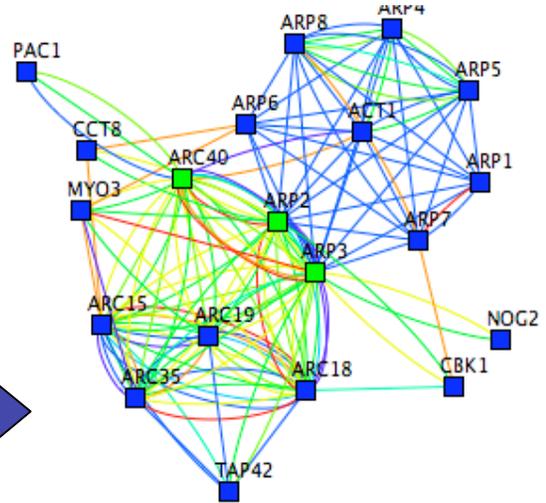
- Must carefully match data sets to avoid errors e.g. different interaction experiments
- Consider data set bias
- Consider binary vs. discrete vs. continuous



Hwang D et al. A data integration methodology for systems biology  
Proc Natl Acad Sci U S A. 2005 Nov 29;102(48):17296-301

# Predicting Gene Function

arp2  
arp3  
arc40



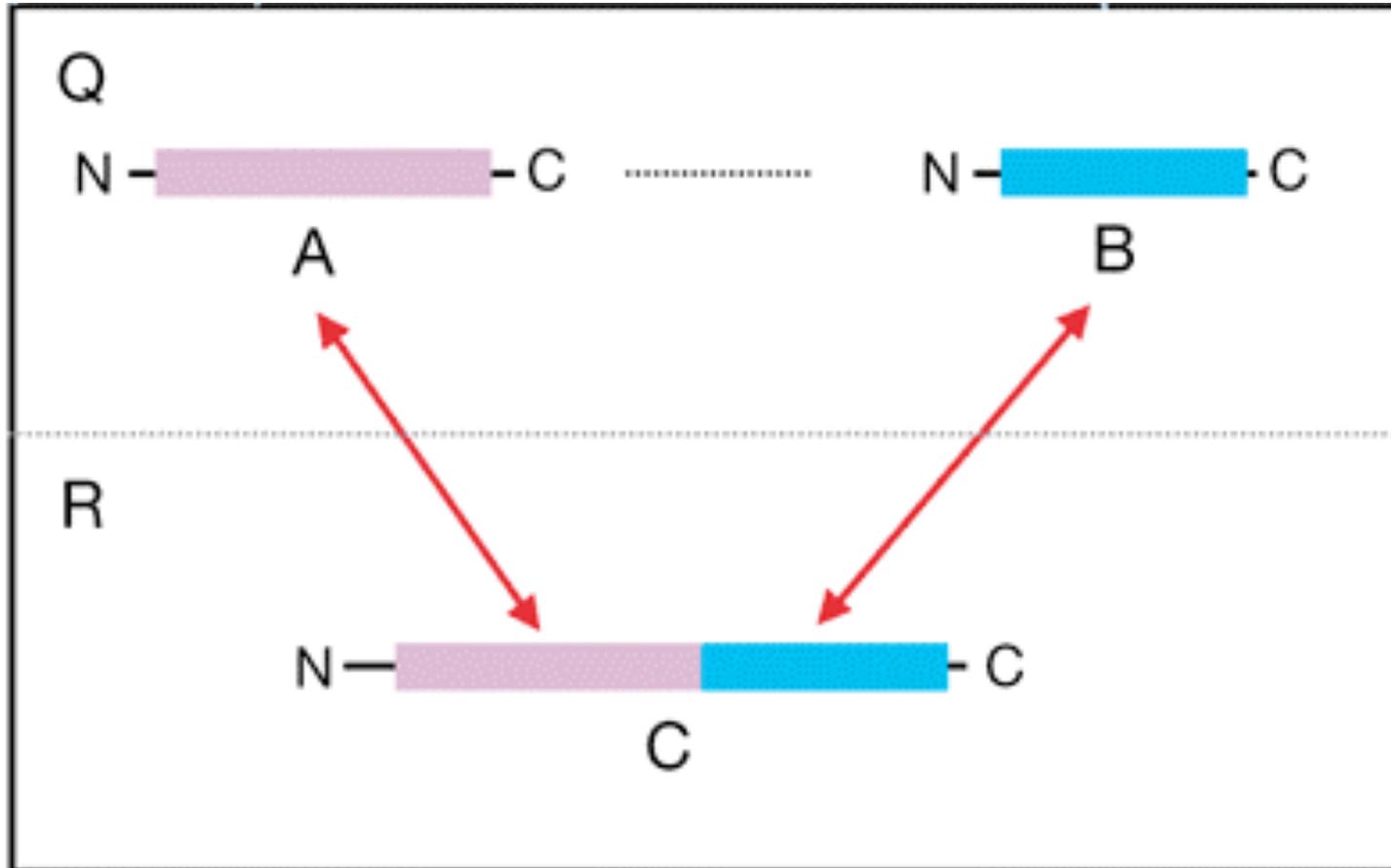
- STRING
  - <http://string.embl.de/>
- bioPIXIE
  - <http://pixie.princeton.edu/pixie/>
- GeneMania
  - <http://genemania.org>

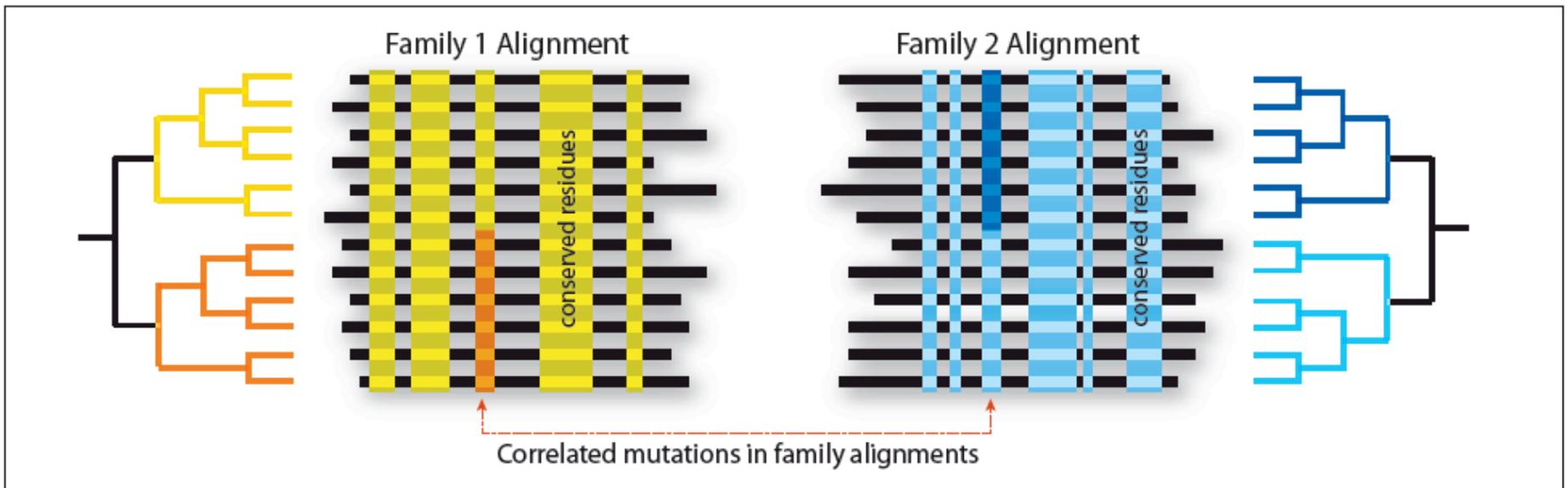
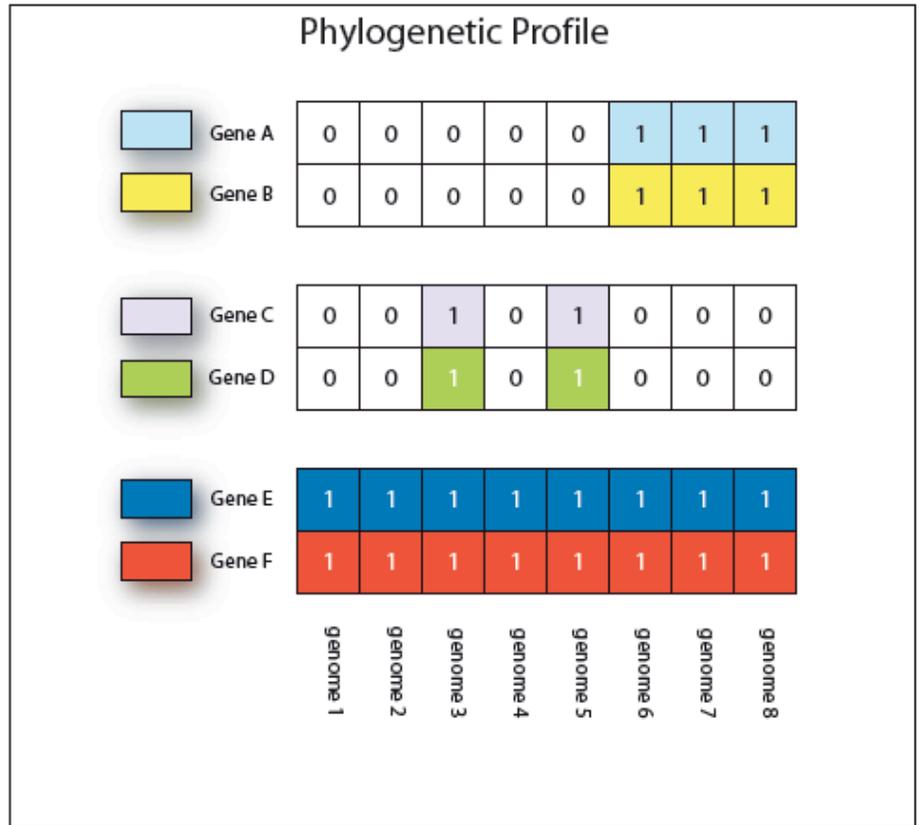
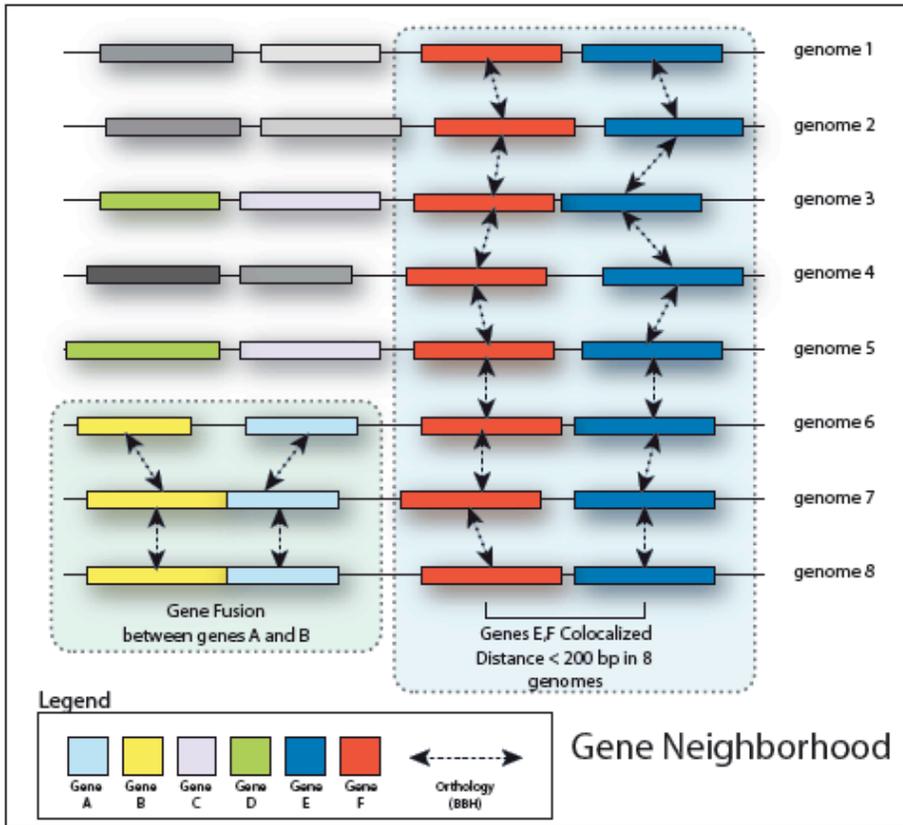
## Top-Scoring Genes

<a href="#">ARC15</a>	0.09026
<a href="#">ARC19</a>	0.08677
<a href="#">ARC35</a>	0.08414
<a href="#">ARC18</a>	0.07793
<a href="#">ARC40</a>	0.03239
<a href="#">ARP8</a>	0.02344
<a href="#">ARP5</a>	0.02293
<a href="#">ARP6</a>	0.02031
<a href="#">TAP42</a>	0.02017
<a href="#">ACT1</a>	0.01854
<a href="#">ARP4</a>	0.01841
<a href="#">ARP1</a>	0.01752
<a href="#">NOG2</a>	0.01676
<a href="#">PAC1</a>	0.01563
<a href="#">ARP7</a>	0.01561
<a href="#">MYO3</a>	0.01551

Fraser AG, Marcotte EM - A probabilistic view of gene function - Nat Genet. 2004 Jun;36(6):559-64

# Gene Fusion





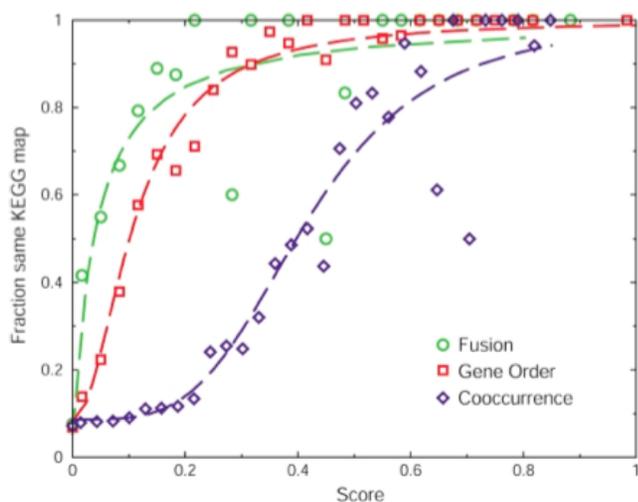
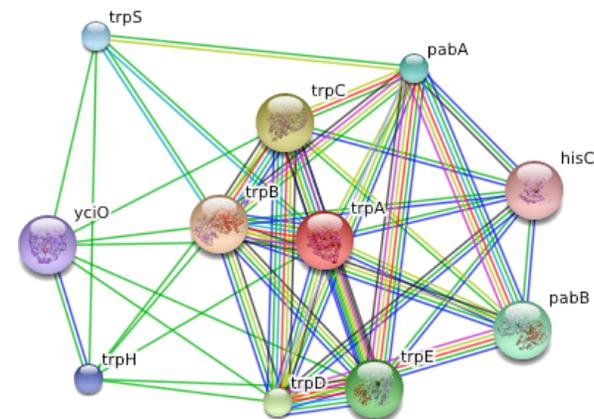
# STRING

## Your Input:

- trpA Tryptophan synthase alpha chain (EC 4.2.1.20) (268 aa) (*Escherichia coli* K12)

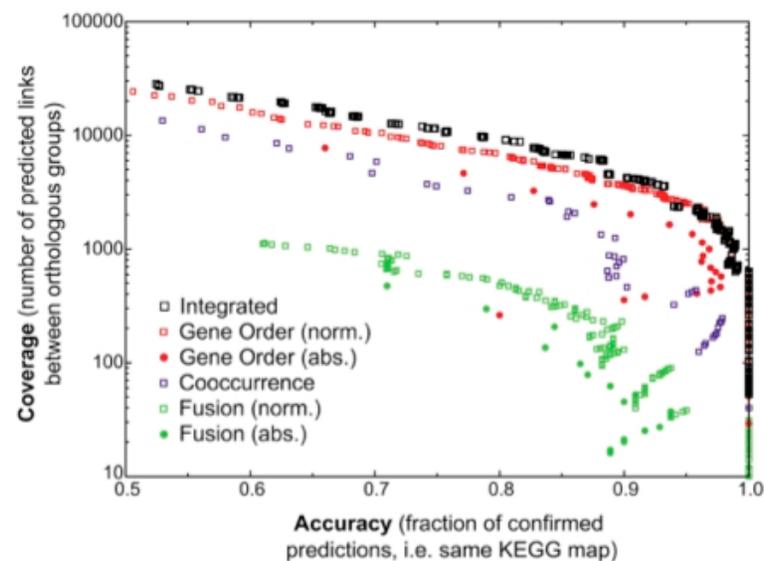
## Predicted Functional Partners:

	Neighborhood	Gene Fusion	Cooccurrence	Experiments	Databases	Textmining	[Homology]	Score
● trpB Tryptophan synthase beta chain (EC 4.2.1.20) (397 aa)	●	●	●	●	●	●	●	0.999
● trpC Tryptophan biosynthesis protein trpCF [Includes- Indole-3-glycerol phosphate sy	●	●	●	●	●	●	●	0.999
● trpD Anthranilate synthase component II (EC 4.1.3.27) [Includes- Glutamine amidotr	●	●	●	●	●	●	●	0.999
● trpE Anthranilate synthase component 1 (EC 4.1.3.27) (Anthranilate synthase compo	●	●	●	●	●	●	●	0.999
● pabB Para-aminobenzoate synthase component 1 (EC 6.3.5.8) (Para- aminobenzoate	●	●	●	●	●	●	●	0.914
● pabA Para-aminobenzoate synthase glutamine amidotransferase component II (EC 6.	●	●	●	●	●	●	●	0.903
● trpS Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan--tRNA ligase) (TrpRS) (	●	●	●	●	●	●	●	0.864
● trpH Protein trpH {UniProtKB/Swiss-Prot-P77766} (293 aa)	●	●	●	●	●	●	●	0.845
● yciO Protein yciO {UniProtKB/Swiss-Prot-P0AFR4} (206 aa)	●	●	●	●	●	●	●	0.841
● hisC Histidinol-phosphate aminotransferase (EC 2.6.1.9) (Imidazole acetol- phosphat	●	●	●	●	●	●	●	0.830

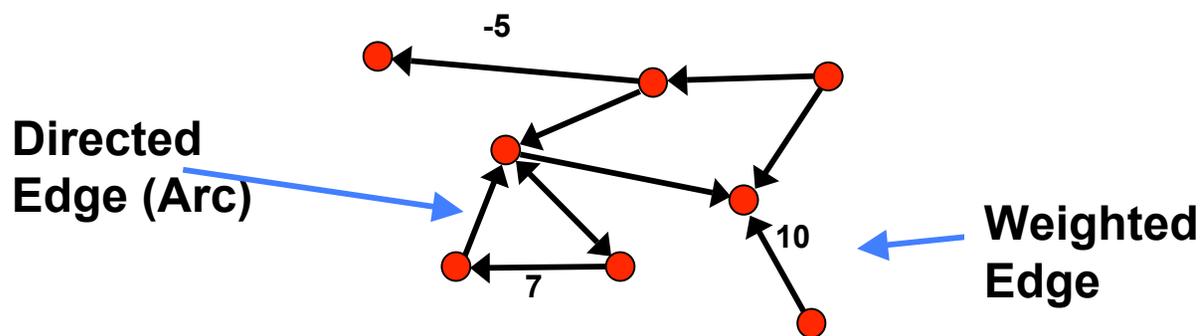
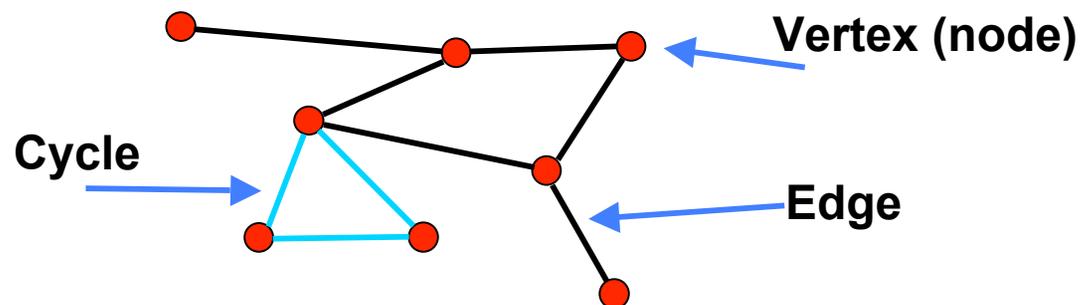


x=score  
a=intercept  
b=cooperativity  
c=half max for x

$$f(x) = a + [(1 - a)x^b / (c^b + x^b)]$$



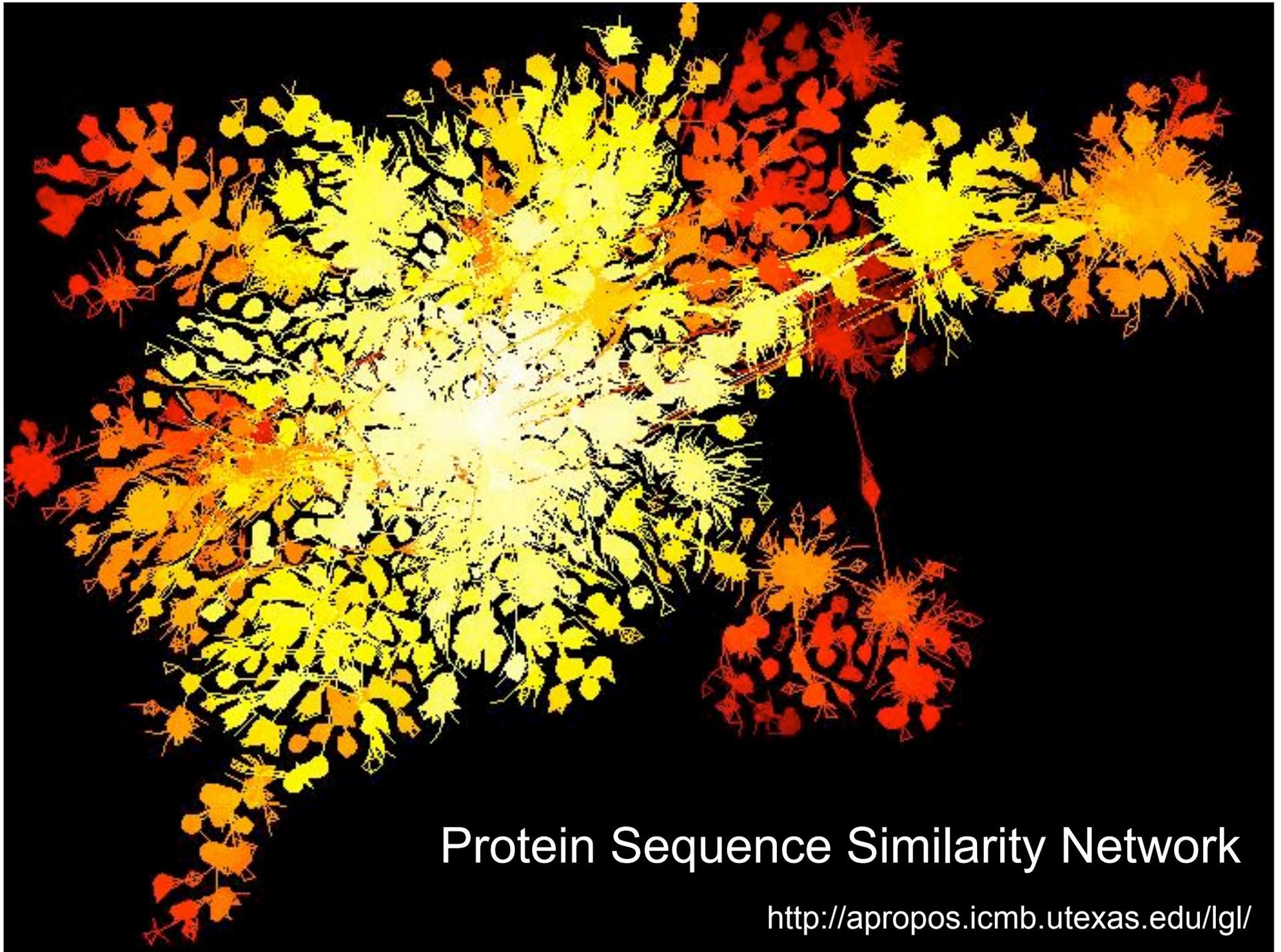
# Graph Theory



We map molecular interaction networks to graphs

# Mapping Biology to a Network

- A simple mapping
  - one compound/node, one interaction/edge
- A more realistic mapping
  - Cell localization, cell cycle, cell type, taxonomy
  - Only represent physiologically relevant interaction networks
- Edges can represent other relationships
- **Critical:** understand the mapping for network analysis



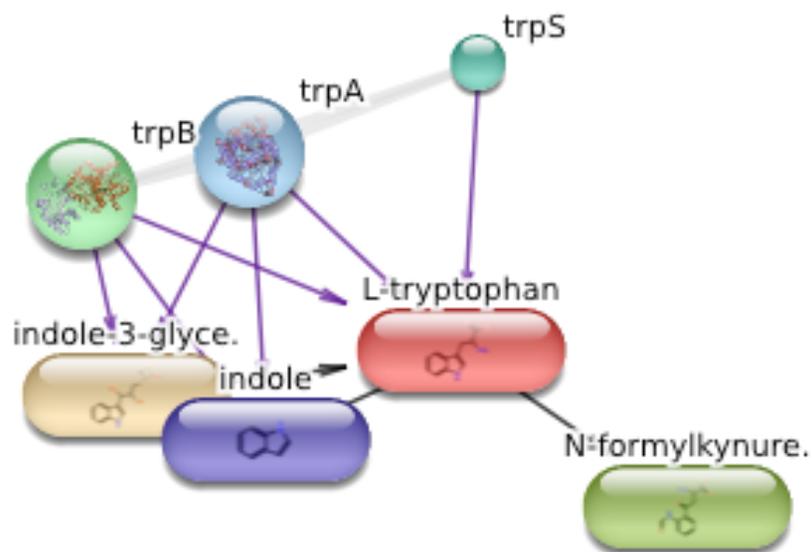
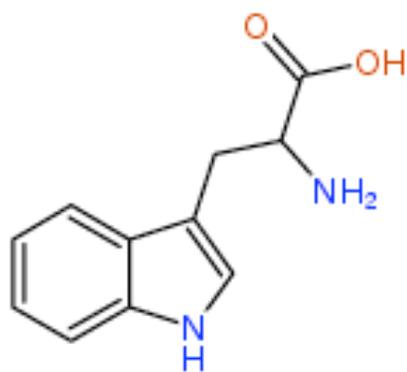
# Protein Sequence Similarity Network

<http://apropos.icmb.utexas.edu/lgl/>

# STITCH

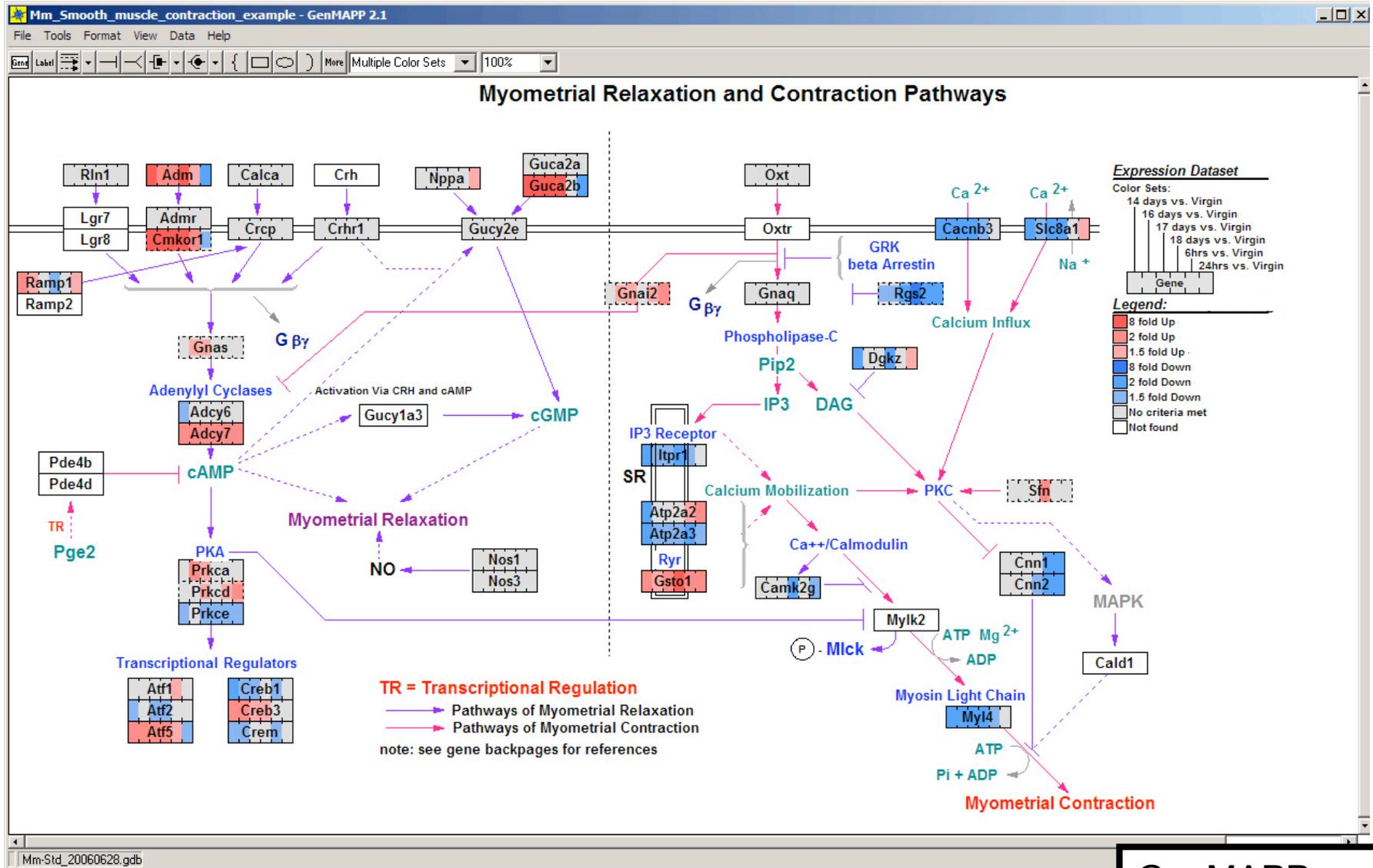
- STITCH: Chemical-Protein Interactions
- <http://stitch.embl.de/>

Interaction network around  
tryptophan:





# Visualizing Time Course Data on Pathways: Multiple Comparison View



http://pathguide.org

Vuk Pavlovic

# Pathguide» the pathway resource list

Home BioPAX cBio MSKCC

## >240 Pathway Databases!

### Navigation

- Protein-Protein Interactions
- Metabolic Pathways
- Signaling Pathways
- Pathway Diagrams
- Transcription Factors / Gene Regulatory Networks
- Protein-Compound Interactions
- Genetic Interaction Networks
- Protein Sequence Focused
- Other

### Search

Organisms  
All

Availability  
All

Standards  
All

Reset Search

### Statistics

Analyze Pathguide

### Contact

Comments, Questions, Suggestions are Always Welcome!

### Complete Listing of All Pathguide Resources

Pathguide contains information about **222** biological pathway resources. Click on a link to go to the resource home page or 'Details' for a description page. Databases that are free and those supporting BioPAX, CellML, PSI-M, or SBML standards are respectively indicated.

If you know of a pathway resource that is not listed here, or have other questions or comments, please [send us an e-mail](#).

Get the Stats  
Detailed Pathguide resource statistics now available

Pathguide Published  
Please cite the [Pathguide](#)

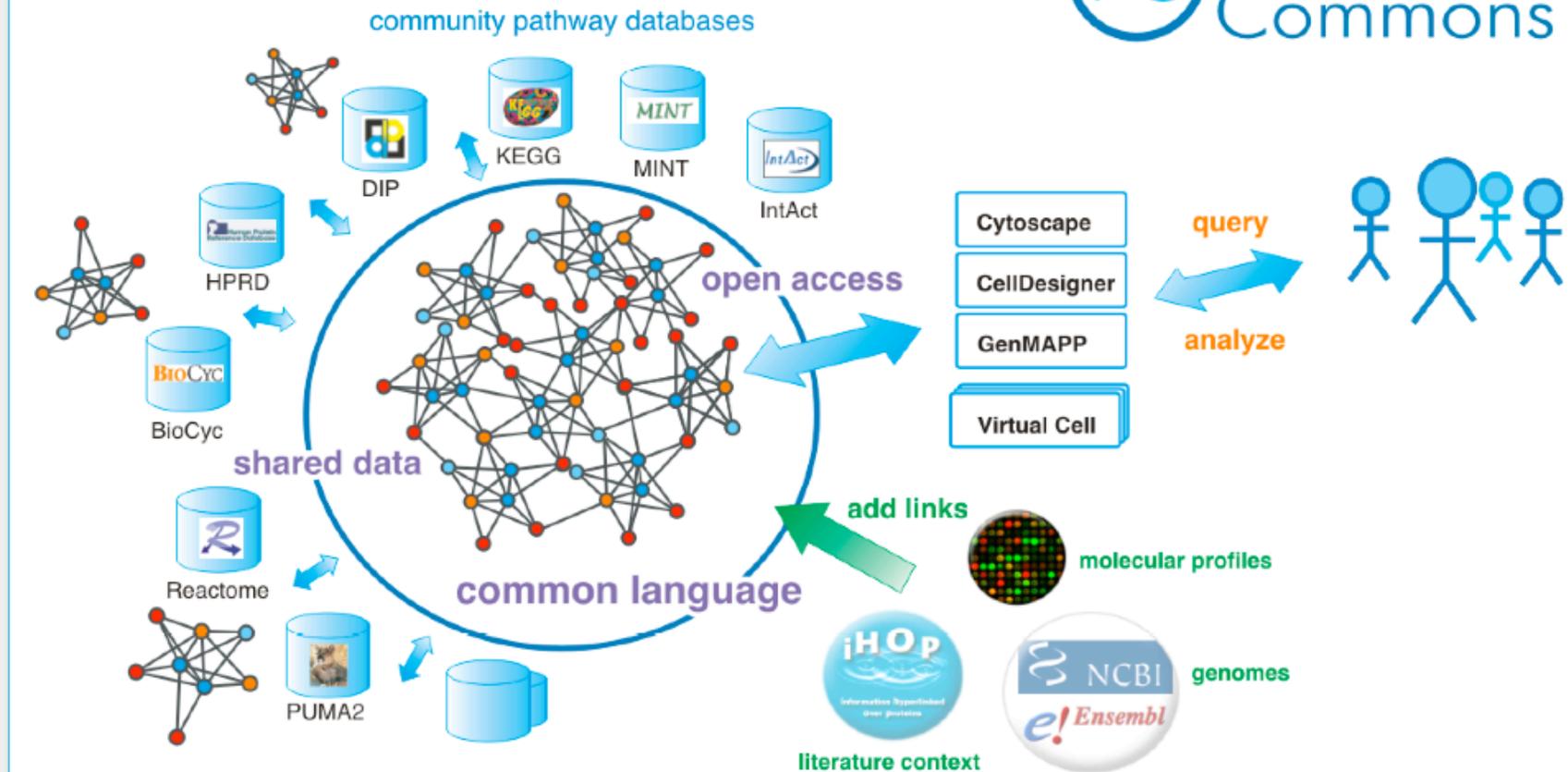
### Protein-Protein Interactions

Database Name (Order: alphabetically   <a href="#">by web popularity</a> )	Full Record	Availability	Standards
3DID - 3D interacting domains	<a href="#">Details</a>	Free	
ABCdb - Archaea and Bacteria ABC transporter database	<a href="#">Details</a>	Free	
AfCS - Alliance for Cellular Signaling Molecule Pages Database	<a href="#">Details</a>	Free	
AllFuse - Functional Associations of Proteins in Complete Genomes	<a href="#">Details</a>	Free	
ASEdb - Alanine Scanning Energetics Database	<a href="#">Details</a>	Free	
ASPD - Artificial Selected Proteins/Peptides Database	<a href="#">Details</a>	?	
BID - Binding Interface Database	<a href="#">Details</a>	Free	
BIND - Biomolecular Interaction Network Database	<a href="#">Details</a>	Free	PSI-MI
BindingDB - The Binding Database	<a href="#">Details</a>	Free	
BioGRID - General Repository for Interaction Datasets	<a href="#">Details</a>		PSI-MI
BRITE - Biomolecular Relations in Information Transmission and Expression	<a href="#">Details</a>	Free	
CA1Neuron - Pathways of the hippocampal CA1 neuron	<a href="#">Details</a>	Free	
Cancer Cell Map - The Cancer Cell Map	<a href="#">Details</a>	Free	BioPAX
CSP - Cytokine Signaling Pathway Database	<a href="#">Details</a>	Free	
CTDB - Calmodulin Target Database	<a href="#">Details</a>	Free	
DDIB - Database of Domain Interactions and Bindings	<a href="#">Details</a>	Free	
DIP - Database of Interacting Proteins	<a href="#">Details</a>		PSI-MI
Doodle - Database of oligomeri			
DopaNet - DopaNet			
DRC - Database of Ribosomal C			
DSM - Dynamic Signaling Maps			
FIMM - Functional Molecular Im			
FusionDB - Prokaryote Gene Fu			

- Varied formats, representation, coverage
- Pathway data extremely difficult to combine and use

# Aim: Convenient Access to Pathway Information

<http://www.pathwaycommons.org>



Facilitate creation and communication of pathway data  
Aggregate pathway data in the public domain  
Provide easy access for pathway analysis

Long term: Converge  
to integrated cell map

http://pathwaycommons.org

Pathway Commons is a convenient point of access to biological pathway information collected from public pathway databases, which you can browse or search. Pathways include biochemical reactions, complex assembly, transport and catalysis events, and physical interactions involving proteins, DNA, RNA, small molecules and complexes. [more...](#)

**Search Pathway Commons:**

Search

To get started, enter a gene name, gene identifier or pathway name in the text box above. For example: [p53](#), [P38398](#) or [mTOR](#).

To restrict your search to specific data sources or specific organisms, update your [global filter settings](#).

**Pathway Commons Quick Stats:**

Number of Pathways:	921
Number of Interactions:	9,924
Number of Physical Entities:	15,515
Number of Organisms:	10

**Biologists:** Browse and search pathways across multiple valuable public pathway databases.

**Computational biologists:** Download an integrated set of pathways in BioPAX format for global analysis.

**Software developers:** Build software on top of Pathway Commons using our soon-to-be released web service API. Download and install the [cPath software](#) to create a local mirror.

Pathway Commons currently contains the following data sources:



[Cancer Cell Map, Release: 1.0](#) [19-May-06]

[Browse](#)



[HumanCyc, Release: 10.5](#) [18-Sep-06]

[Browse](#)



[NCI / Nature Pathway Interaction Database](#)

[01-Jan-07]

[Browse](#)



[Reactome, Release: 19](#) [16-Nov-06]

[Browse](#)

Searched for: **p53**

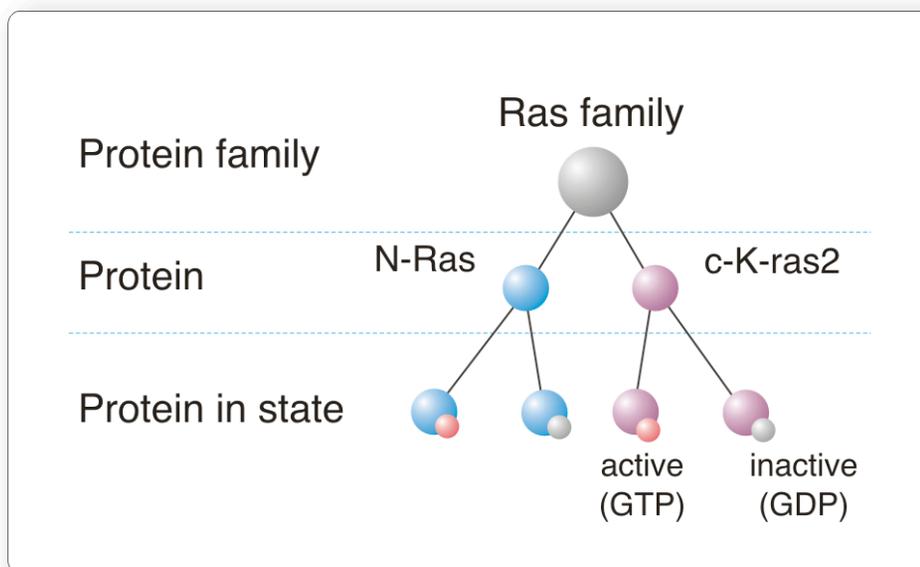
Pathway Commons completed your search for "p53" and found **22** relevant records:

Narrow Results by Type:	Showing Results 1 - 10 of 22   Next 10
<ul style="list-style-type: none"> <li>All Types (45)</li> <li>Pathway (22) ▾</li> <li>Protein (23)</li> </ul>	<p><b>Pathway: Transcriptional activation of p53 responsive genes</b> <span style="float:right">-</span></p> <p><b>Summary:</b></p> <p><b>p53</b> causes G1 arrest by inducing the expression of a cell cycle inhibitor, p21 (El-Deiry et al, 1993; Harper et al, 1993; Xiong et al, 1993). P21 binds and inactivates Cyclin-Cdk complexes that mediate G1/S progression, resulting in lack of phosphorylation of Rb, E2F sequestration and cell cycle arrest at the G1/S transition. Mice with a homozygous deletion of p21 gene are deficient in their ability to undergo a G1/S arrest in response to DNA damage (Deng et al, 1995).</p> <p><b>Data Sources:</b></p> <ul style="list-style-type: none"> <li>Reactome</li> <li>... <b>p53</b> causes G1 arrest by inducing the expression of a cell cycle inhibitor, p21 (El-Deiry et al, 1993; Harper et al, 1993; Xiong et al, 1993).</li> </ul>
<p>Narrow Results by Data Source:</p> <ul style="list-style-type: none"> <li>All Data Sources (22) ▾</li> <li>Cancer Cell Map (2)</li> <li>NCI / Nature Pathway Interaction Database (3)</li> <li>Reactome (17)</li> </ul> <p><a href="#">[Update Filter Settings]</a></p>	<p><b>Pathway: Stabilization of p53</b> <span style="float:right">+</span></p> <ul style="list-style-type: none"> <li>... ATM also regulates the phosphorylation of <b>p53</b> at other sites, especially Ser-20, by activating other serine/threonine kinases in response to IR (Chehab et al, 2000 ...</li> </ul>
	<p><b>Pathway: p53-Dependent G1 DNA Damage Response</b> <span style="float:right">+</span></p> <ul style="list-style-type: none"> <li>Most of the damage-induced modifications of <b>p53</b> are dependent on the ATM kinase. ... The first link between ATM and <b>p53</b> was predicted based on the earlier studies that showed that AT cells exhibit a reduced and delayed induction of <b>p53</b> following exposure to IR (Kastan et al, 1992 and Khanna and Lavin, 1993). ... Under normal conditions, <b>p53</b> is a short-lived protein ...</li> </ul>
	<p><b>Pathway: p53-Dependent G1/S DNA damage checkpoint</b> <span style="float:right">+</span></p> <ul style="list-style-type: none"> <li>The arrest at G1/S checkpoint is mediated by the action of a widely known tumor suppressor protein, <b>p53</b>. ... Loss of <b>p53</b> functions, as a result of mutations in cancer prevent the G1/S checkpoint (Kuerbitz et al, 1992). ... <b>P53</b> is rapidly induced in response to damaged DNA.</li> </ul>
	<p><b>Pathway: p53-Independent G1/S DNA damage checkpoint</b> <span style="float:right">+</span></p> <ul style="list-style-type: none"> <li>The G1 arrest induced by DNA damage has been ascribed to the transcription factor and tumor suppressor protein <b>p53</b>.</li> </ul>
	<p><b>Pathway: G1/S DNA Damage Checkpoints</b> <span style="float:right">+</span></p> <ul style="list-style-type: none"> <li>In the G1 phase there are two types of DNA damage responses, the p53-dependent and the p53-independent pathways. ... The p53-dependent responses inhibit CDKs through the up-regulation of genes encoding CKIs mediated by the <b>p53</b> protein, whereas the p53-independent mechanisms inhibit CDKs through the inhibitory T14Y15 phosphorylation of Cdk2.</li> </ul>
	<p><b>Pathway: Cell Cycle Checkpoints</b> <span style="float:right">+</span></p>

<http://pathwaycommons.org>

# Towards an Integrated Cell Map

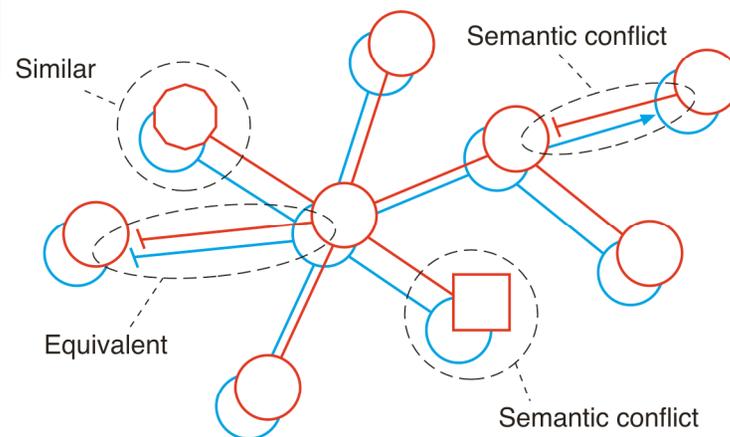
- Semantic pathway integration is difficult



## Physical entities

Determining equivalent entities is critical

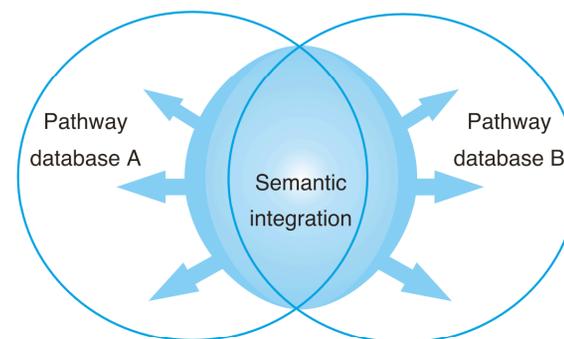
## Relationships



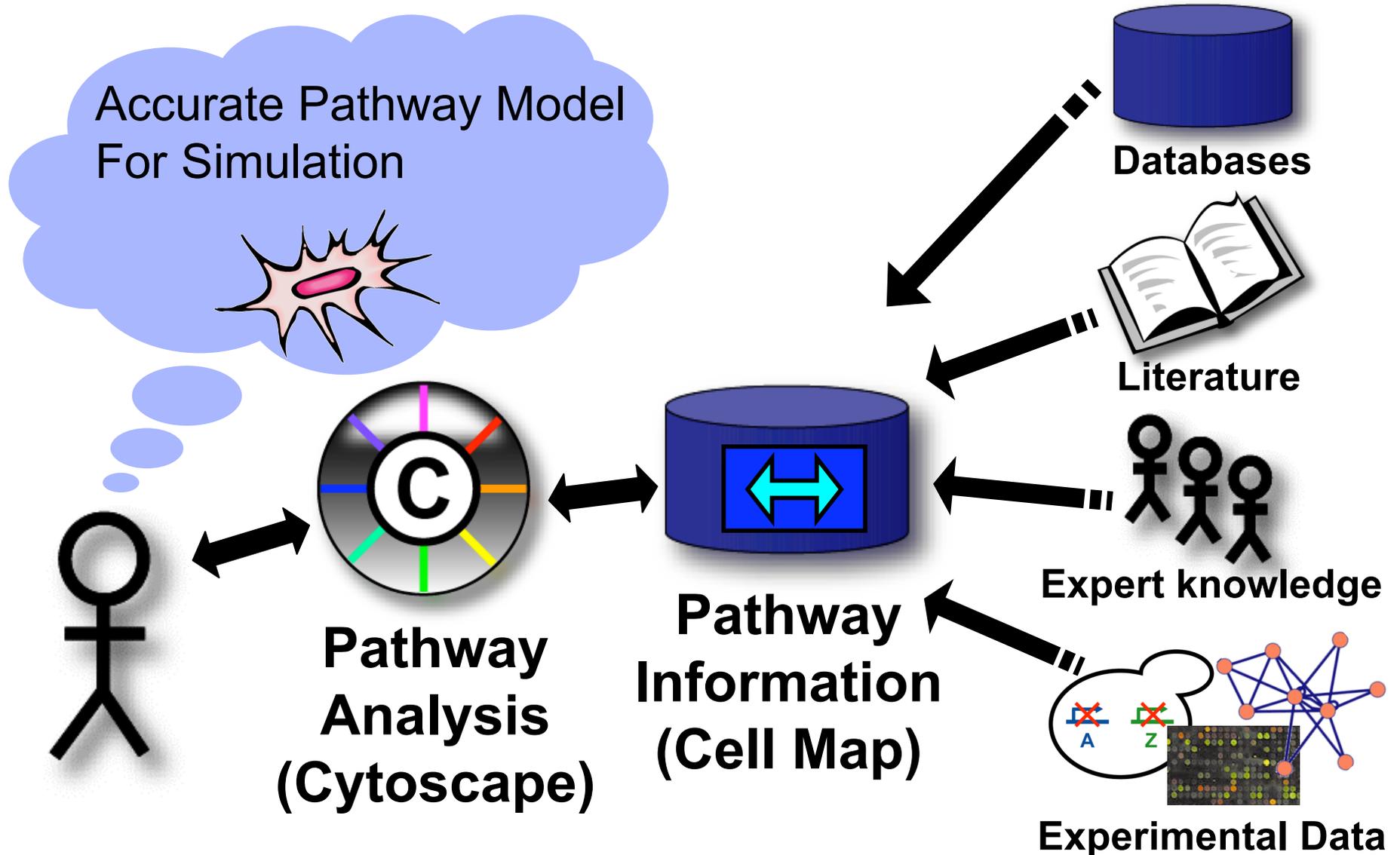
# Practical Semantic Integration

- Minimize errors
  - Integrate only where possible with high accuracy
  - Detect and flag conflicts, errors for users, no revision
  - Promote best-practices to minimize future errors
  - Interaction confidence algorithms
  - Validation software
  - Allow users to filter and select trusted sources
- Converge to standard representation
  - Community process

Doable: hundreds of curators globally in >200 databases (GDP) - make it more efficient



# Using Pathway Information

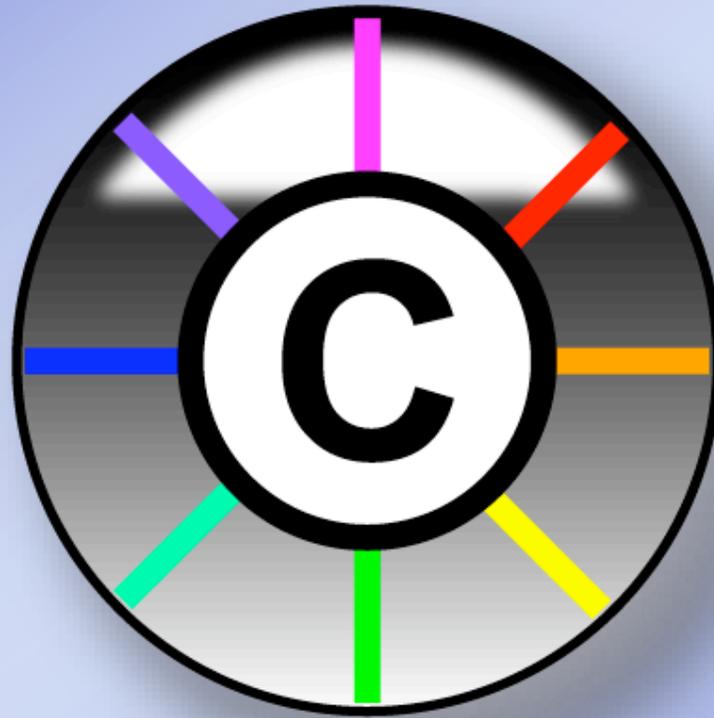




# Cytoscape



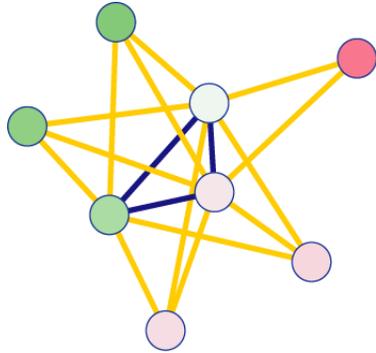
<http://cytoscape.org>



**Agilent Technologies**



# Cytoscape - Network Visualization and Analysis



<http://cytoscape.org>

- Freely-available (open-source, java) software
- Visualizing biological networks (e.g. molecular interaction networks)
- Analyzing networks with gene expression profiles and other cell state data

Other software: Osprey, BioLayout, VisANT, Navigator, PIMWalker, ProViz

Cytoscape Desktop (Session: galFiltered.cys)

Control Panel

Network

Network Nodes Edges  
galFiltered.sif 331(4) 362(0)

galFiltered.sif

Data Panel

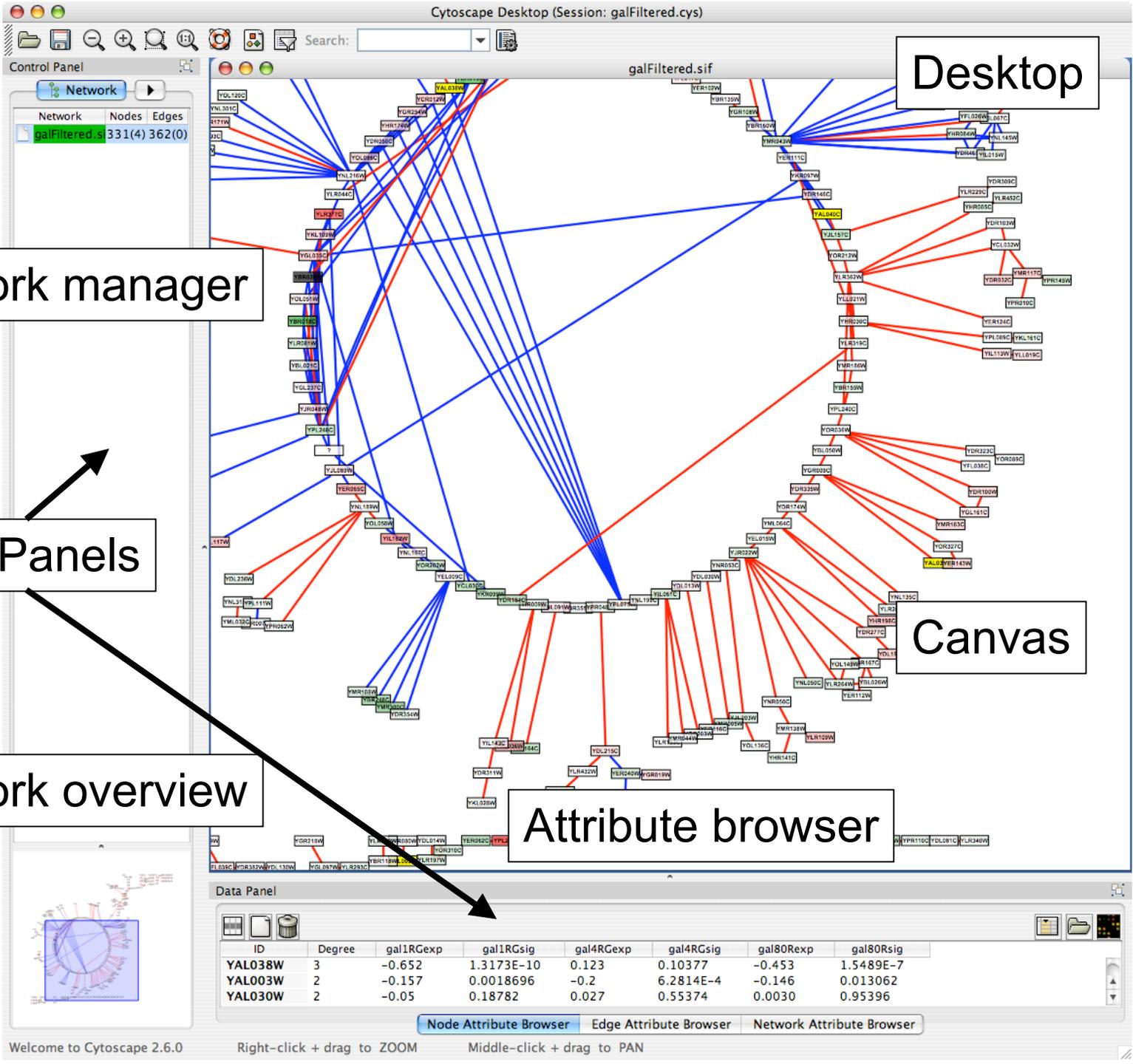
ID	Degree	gal1RGexp	gal1RGsig	gal4RGexp	gal4RGsig	gal80Rexp	gal80Rsig
YAL038W	3	-0.652	1.3173E-10	0.123	0.10377	-0.453	1.5489E-7
YAL003W	2	-0.157	0.0018696	-0.2	6.2814E-4	-0.146	0.013062
YAL030W	2	-0.05	0.18782	0.027	0.55374	0.0030	0.95396

Node Attribute Browser | Edge Attribute Browser | Network Attribute Browser

Welcome to Cytoscape 2.6.0

Right-click + drag to ZOOM

Middle-click + drag to PAN



Desktop

Network manager

CytoPanels

Network overview

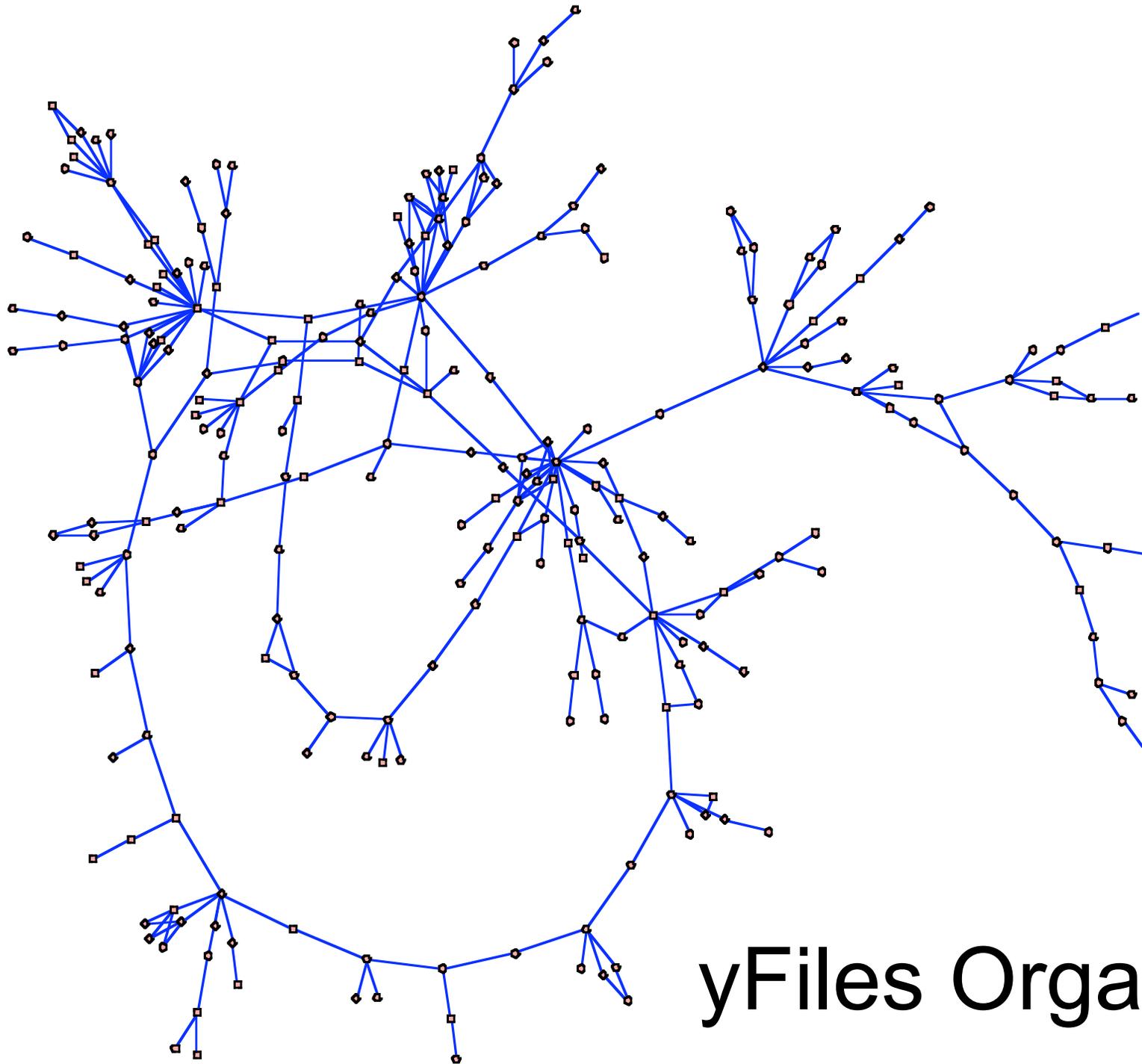
Attribute browser

Canvas

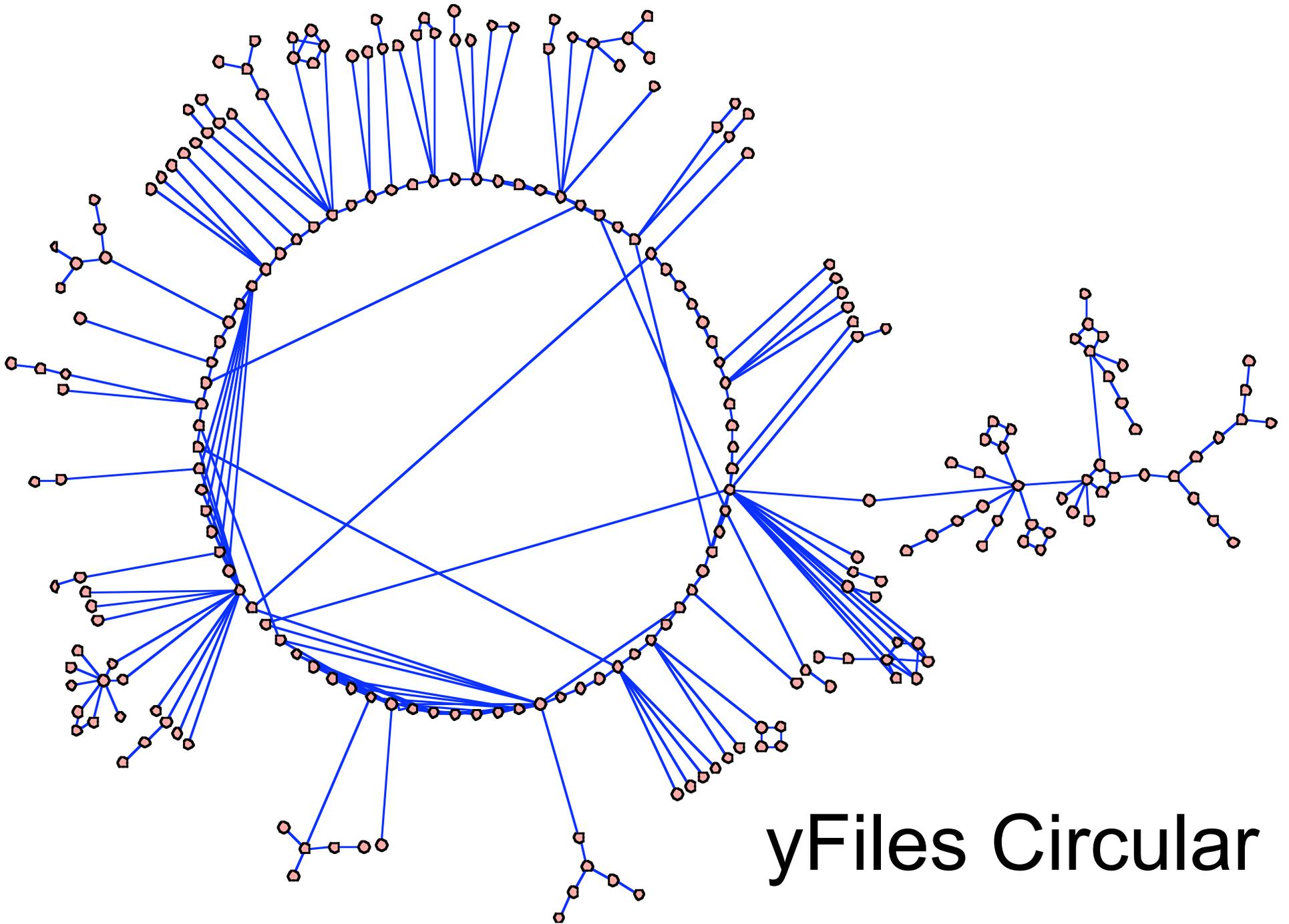
Data Panel

ID	Degree	gal1RGexp	gal1RGsig	gal4RGexp	gal4RGsig	gal80Rexp	gal80Rsig
YAL038W	3	-0.652	1.3173E-10	0.123	0.10377	-0.453	1.5489E-7
YAL003W	2	-0.157	0.0018696	-0.2	6.2814E-4	-0.146	0.013062
YAL030W	2	-0.05	0.18782	0.027	0.55374	0.0030	0.95396

Node Attribute Browser | Edge Attribute Browser | Network Attribute Browser



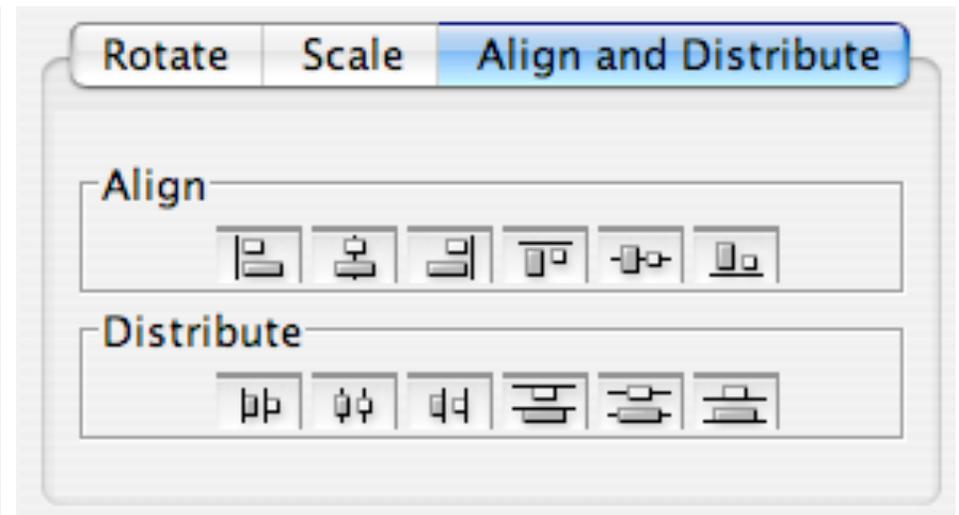
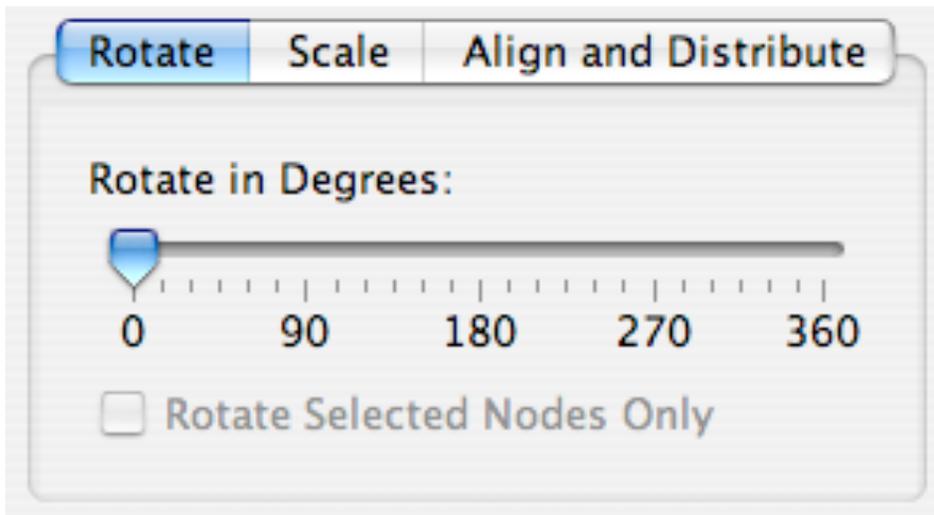
yFiles Organic



yFiles Circular

# Network Layout

- 15 algorithms available through plugins
- Demo: Move, zoom/pan, rotate, scale, align



# Create subnetwork

The screenshot shows the Cytoscape Desktop interface. The main window displays a network diagram with a central hub node (yellow square) and several peripheral nodes (triangles) connected by edges. The 'File' menu is open, and the 'New' submenu is active, showing the option 'From selected nodes, all edges' selected. Below the main window, there are two CytoPanel panels. CytoPanel 5 contains 'Align and Distribute' options. CytoPanel 2 contains a 'Node Attribute Browser' showing a list of node IDs: YFL026W, YDR461W, and YNL145W. The status bar at the bottom indicates 'Welcome to Cytoscape 2.4.1' and provides instructions for zooming and panning.

Cytoscape Desktop (Session: galFiltered.cys)

File Edit View Select Layout Plugins Help

New Network From selected nodes, all edges ^N  
Open ^O Session From selected nodes, selected edges ^⇧N  
Save ^S Clone current network  
Save As... ^⇧S Network Editor Empty Network  
Import Construct network using cPath...  
Export network Nodes Edges  
Print... ^P 331(7) 362(20)  
Quit ^Q

CytoPanel 5

Rotate Scale Align and Distribute

Align

Distribute

CytoPanel 2

ID  
YFL026W  
YDR461W  
YNL145W

Node Attribute Browser Edge Attribute Browser

Welcome to Cytoscape 2.4.1 Right-click + drag to ZOOM Middle-click + drag to PAN

# Create subnetwork

Cytoscape Desktop (Session: galFiltered.cys)

File Edit View Select Layout Plugins Help

default Search:

CytoPanel 1

Network	Nodes	Edges
galFiltered.sif	331(7)	362(20)
galFiltered.sif--	7(0)	7(0)

CytoPanel 5

Rotate Scale **Align and Distribute**

Align

Distribute

CytoPanel 2

ID

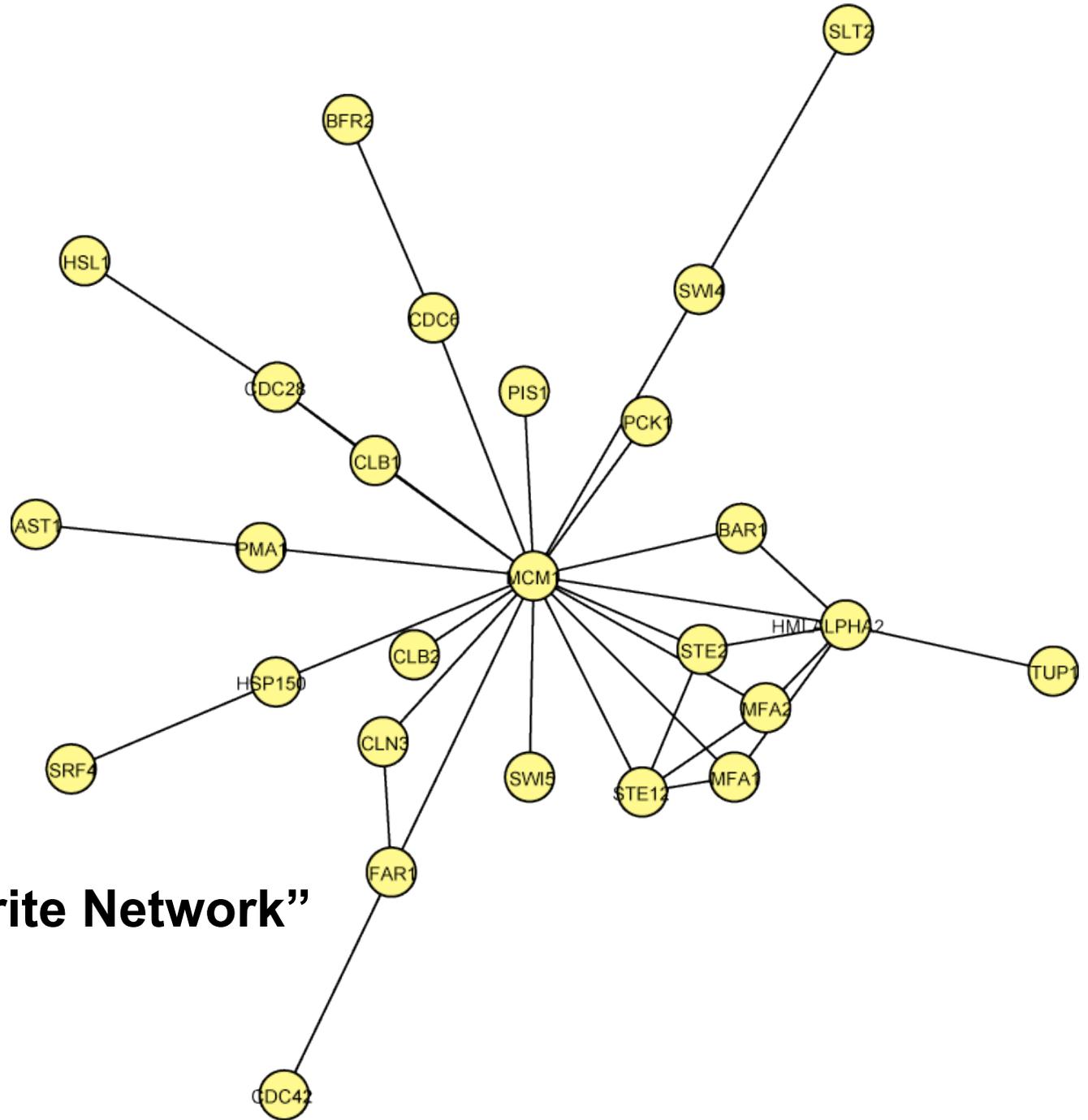
Node Attribute Browser Edge Attribute Browser

Welcome to Cytoscape 2.4.1 Right-click + drag to ZOOM Middle-click + drag to PAN

# Visual Style

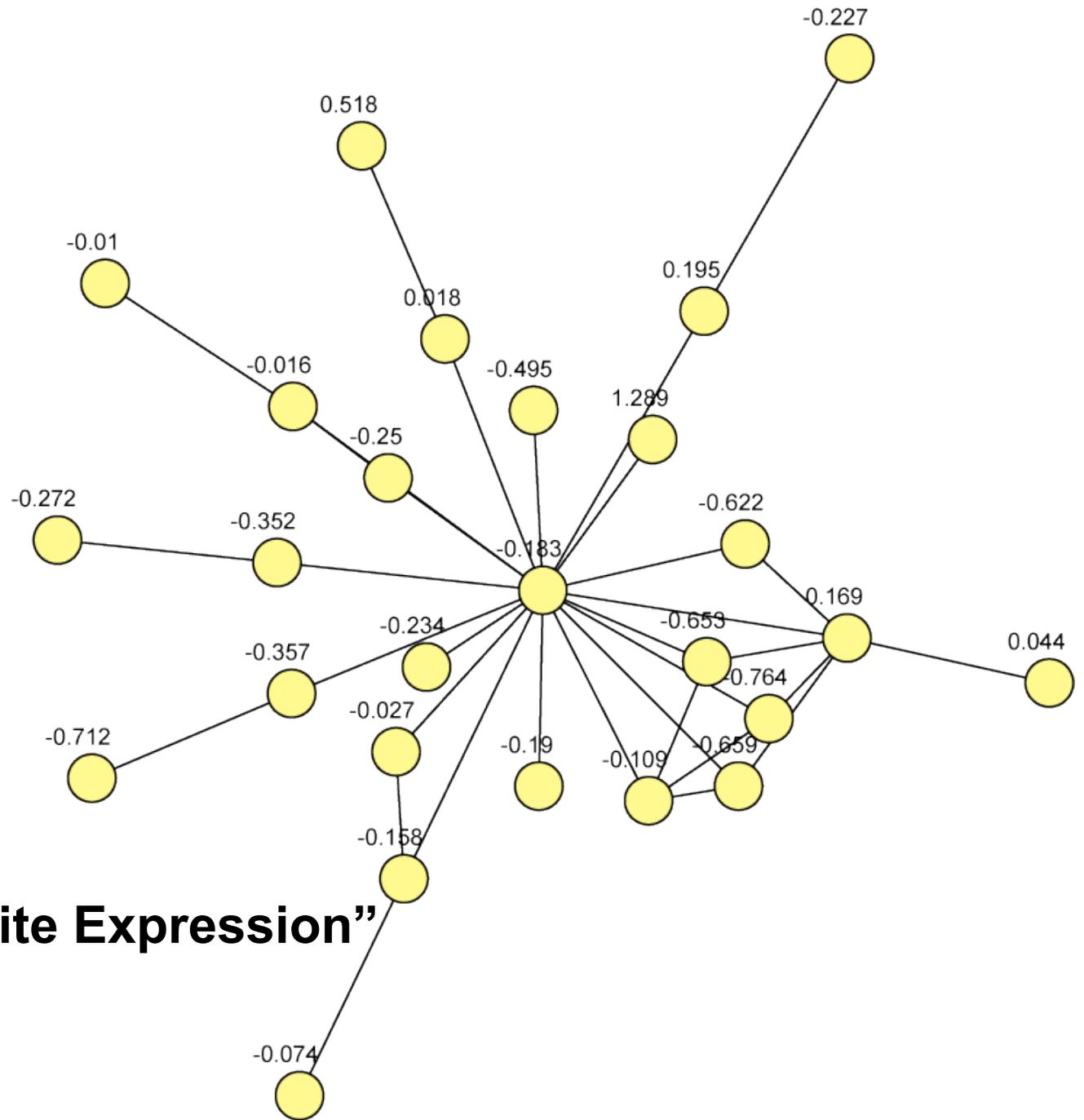
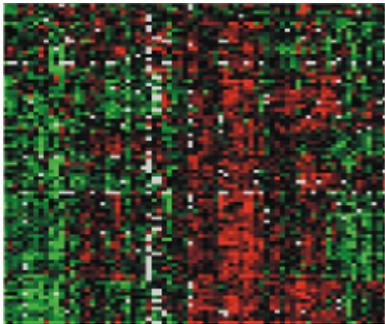
- Customized views of experimental data in a network context
- Network has node and edge attributes
  - E.g. expression data, GO function, interaction type
- Mapped to visual attributes
  - E.g. node/edge size, shape, colour...
- E.g. Visualize gene expression data as node colour gradient on the network

# Visual Style



Load “Your Favorite Network”

# Visual Style



**Load “Your Favorite Expression”  
Dataset**

# Visual Style

The image shows a screenshot of the Cytoscape Desktop application (Session: galFiltered.cys). The main window displays a network graph with nodes and edges. The nodes are colored based on their expression values, ranging from green (low) to red (high). A dialog box titled "Gradient Editor for Node Color" is open, showing a "Continuous Mapping for Node Color" with a color gradient from green to red. The gradient is labeled with values: -2.4060 (Min), -0.00000, and 1.2240 (Max). The mapping is for the property "gal4RGexp". The dialog also shows a "Range Setting" of 0 and buttons for "Min/Max", "Add", and "Delete".

Control Panel

Network VizMapper™ Editor Filters

Current Visual Style: Sample3

Defaults

Source Target

Visual Mapping Browser

- Edge Visual Mapping
  - Edge Color: interaction
- Node Visual Mapping
  - Node Tooltip: gal4RGexp
  - Node Label: ID
  - Node Color: gal4RGexp
    - Mapping Type: Continuous Mapping
    - Graphical View:  -2.41 1.22
- Unused Properties
  - Node Border Color: Double-Click to create...
  - Node Shape: Double-Click to create...
  - Node Width: Double-Click to create...
  - Node Height: Double-Click to create...

galFiltered.sif

Gradient Editor for Node Color

Continuous Mapping for Node Color

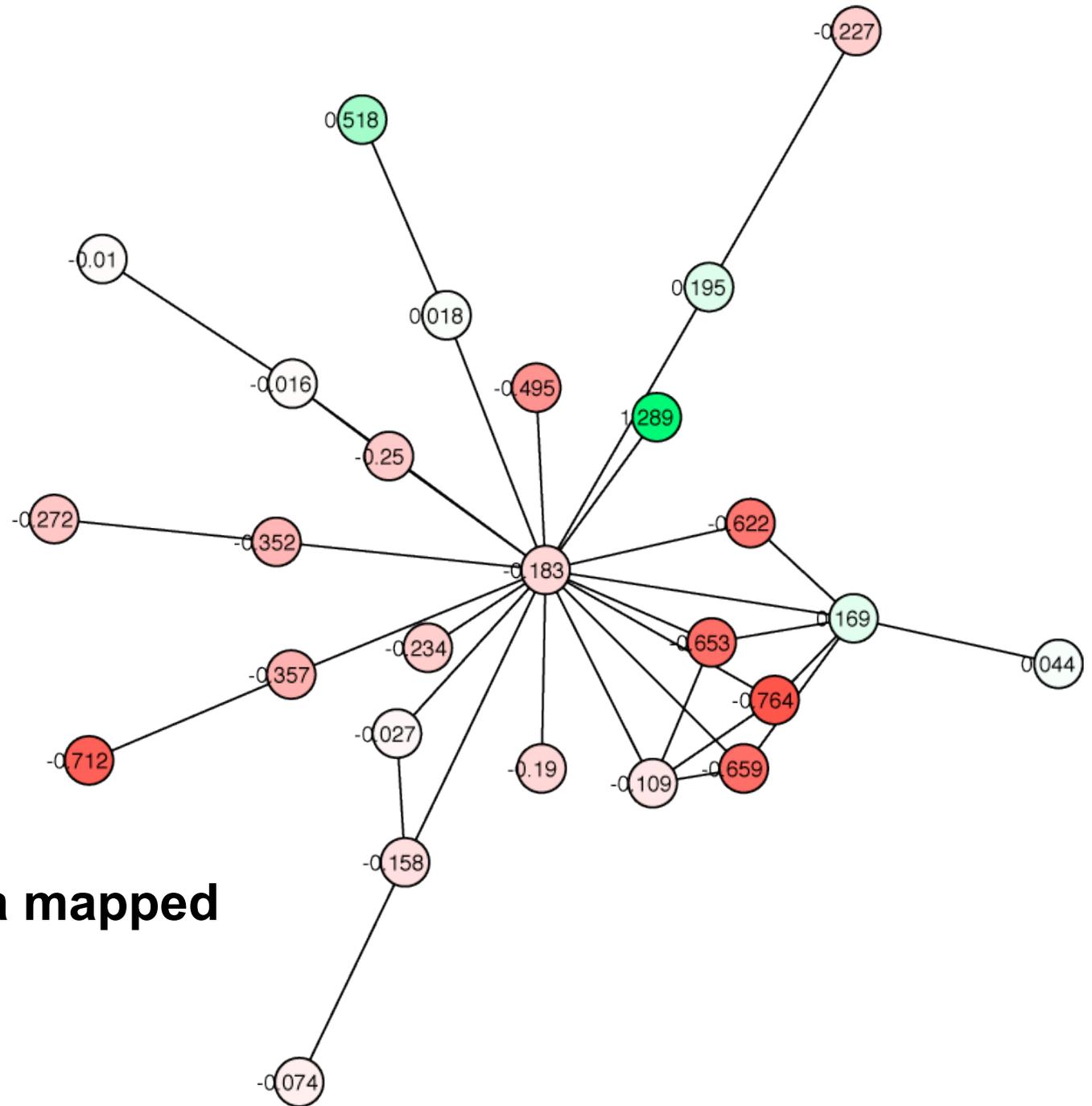
Min = -2.406 gal4RGexp Max = 1.224

Range Setting: 0

Min/Max Add Delete

Map expression values to node colours using a continuous mapper

# Visual Style



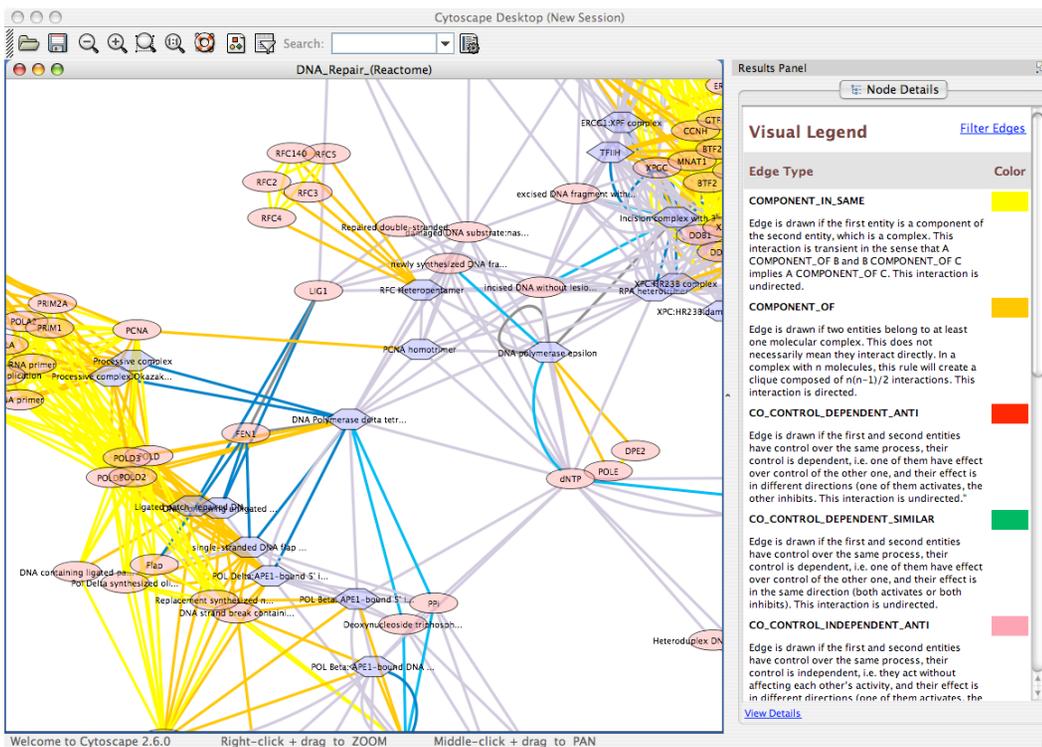
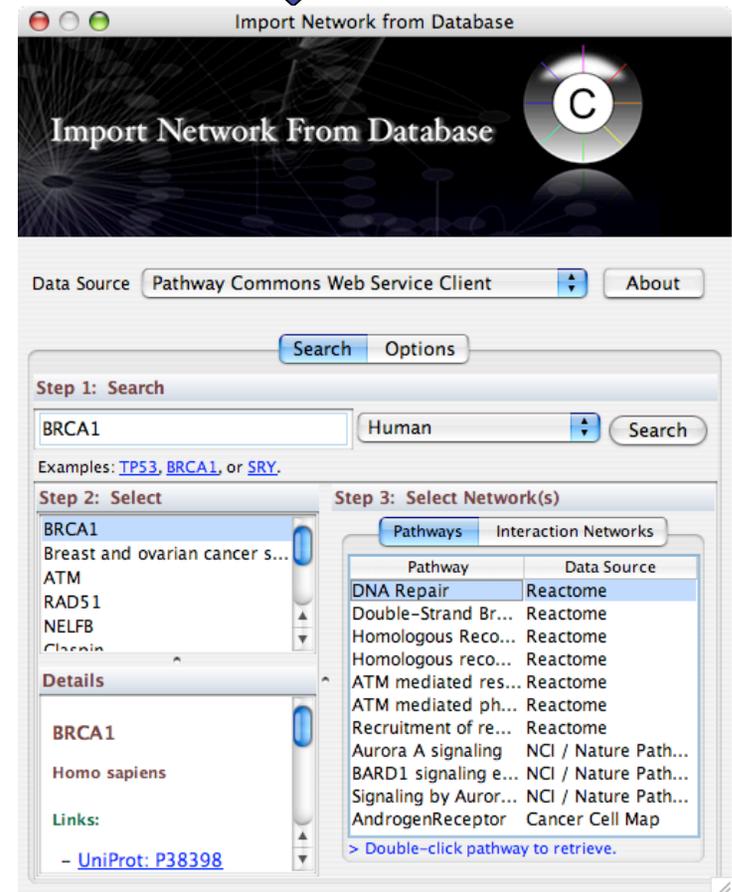
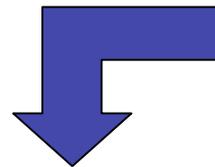
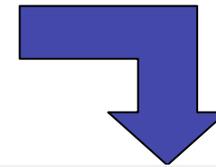
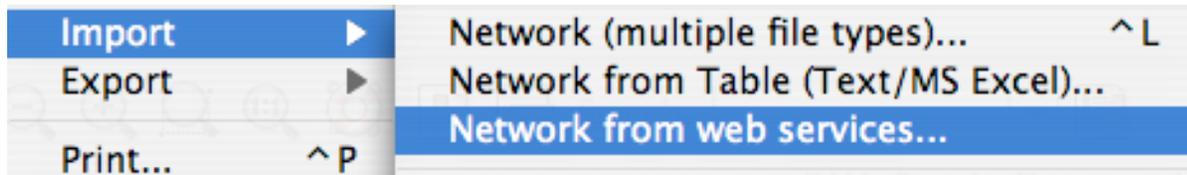
**Expression data mapped  
to node colours**

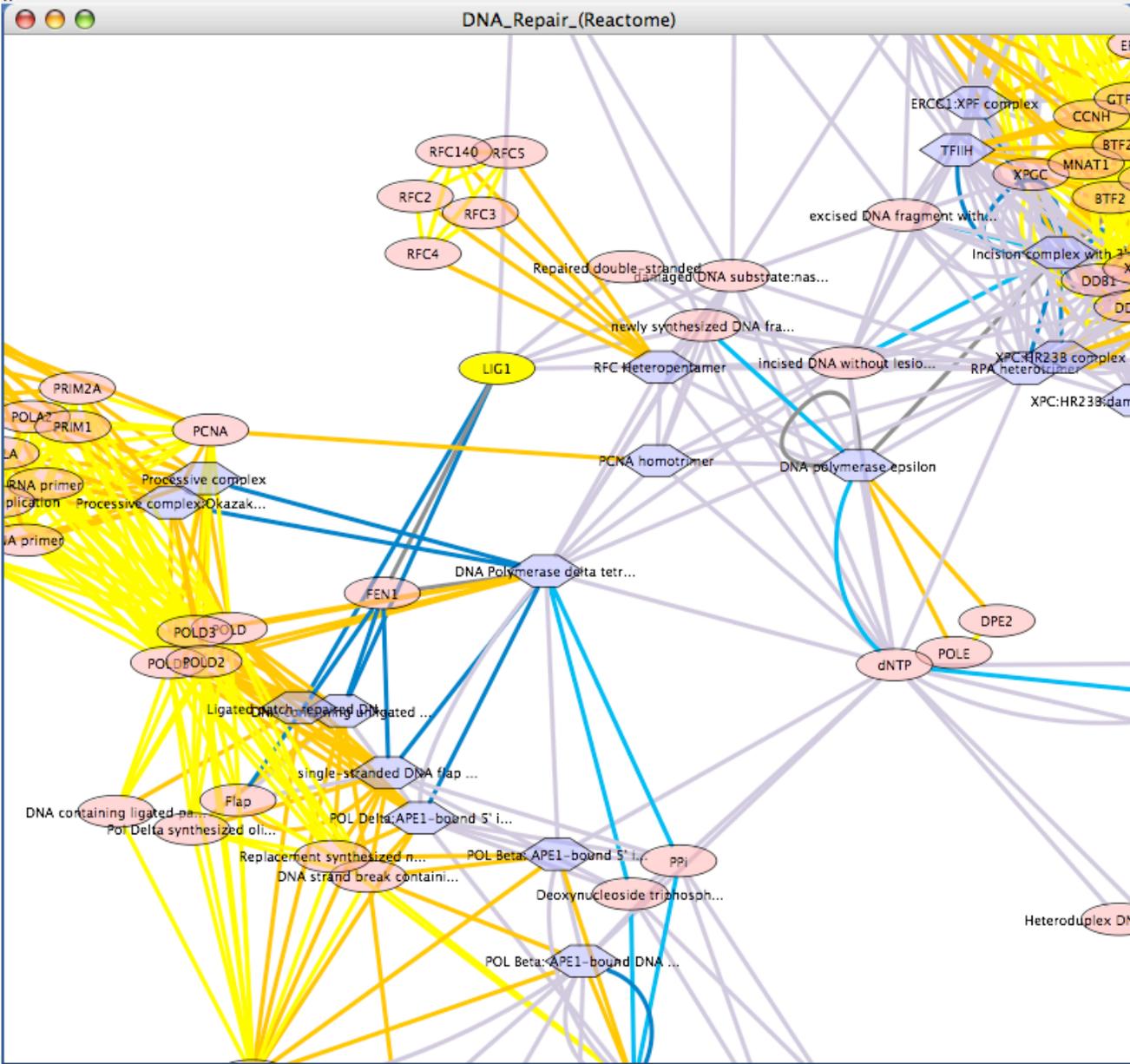


# Analyzing gene expression data in a network context

- Input
  - Gene expression data
  - Network data
- Output
  - Visual diagram of expression data on network
  - Active network regions
- Outline
  - Where to find network data?
    - Interaction database (cPath)
    - Literature associations via text mining
  - Load expression data
  - Identify active pathways

# Interaction Database Search





Node Details

**LIG1**

Protein

Homo sapiens

[Pathway Commons: 6311](#)

Synonyms:

- LIG1

Links:

- [UNIPROT: P18858](#)
- [UNIPROT: Q32P23](#)
- [REF\\_SEQ: NP\\_000225](#)
- [Search iHOP](#)

[Visual Legend](#)



# Text Mining

- Computationally extract gene relationships from text, usually PubMed abstracts
- Literature search tool, lots of network data
- BUT not perfect
  - Problems recognizing gene names
  - Natural language processing not perfect
- Agilent Literature Search Cytoscape plugin
- Others: E.g. iHOP
  - [www.ihop-net.org/UniPub/iHOP/](http://www.ihop-net.org/UniPub/iHOP/)

**Agilent Literature Search 1.0.4**

Edit View Help

**Terms**  
 CSF2RB  
 EDN1  
 EGFR  
 LMNA  
 PDK2  
 TRAF1  
 WBSR14

**Context**  
 atherosclerosis

**Match Controls**  
 Max Engine Matches: 10 Organism: Homo sapiens

**Query Controls**  
 Use Aliases:  Use Context:

**Extraction Controls**  
 Interaction Lexicon: limited

**Query Editor**  
 (((csf2rb OR if5rb OR cd131 OR cdw131 OR il3rb)) AND atherosclerosis  
 ((edn1 OR et1)) AND atherosclerosis  
 ((egfr OR mena OR erbb OR erbb1)) AND atherosclerosis  
 ((lmna OR lmnc OR cmt2b1 OR fpl OR lfp OR hgps OR emd2 OR ldp1 OR lmn1 OR fpld)) AND atherosclerosis  
 (PDK2) AND atherosclerosis  
 ((traf1 OR mgc:10353 OR ebi6)) AND atherosclerosis  
 ((wbscr14 OR ws-bhlh OR chrebp OR mondob OR mio)) AND atherosclerosis

**Query Matches**



**Cytoscape Desktop**

File Edit Data Select Layout Visualization Plugins Help Filters

Network Nodes Edges  
 1 46(0) 77(0)

Nodes: 46 (0 selected) Edges: 77 (0 selected)



Use Aliases:  Use Context:  Interaction Lexicon: limited

**Query Editor**  
 (((csf2rb OR if5rb OR cd131 OR cdw131 OR il3rb)) AND atherosclerosis  
 (CRKL) AND atherosclerosis  
 (((csf2rb OR if5rb OR cd131 OR cdw131 OR il3rb)) AND atherosclerosis  
 ((edn1 OR et1)) AND atherosclerosis  
 ((egfr OR mena OR erbb OR erbb1)) AND atherosclerosis  
 ((lmna OR lmnc OR cmt2b1 OR fpl OR lfp OR hgps OR emd2 OR ldp1 OR lmn1 OR fpld)) AND atherosclerosis  
 (PDK2) AND atherosclerosis  
 ((traf1 OR mgc:10353 OR ebi6)) AND atherosclerosis  
 ((wbscr14 OR ws-bhlh OR chrebp OR mondob OR mio)) AND atherosclerosis

**Query Matches**

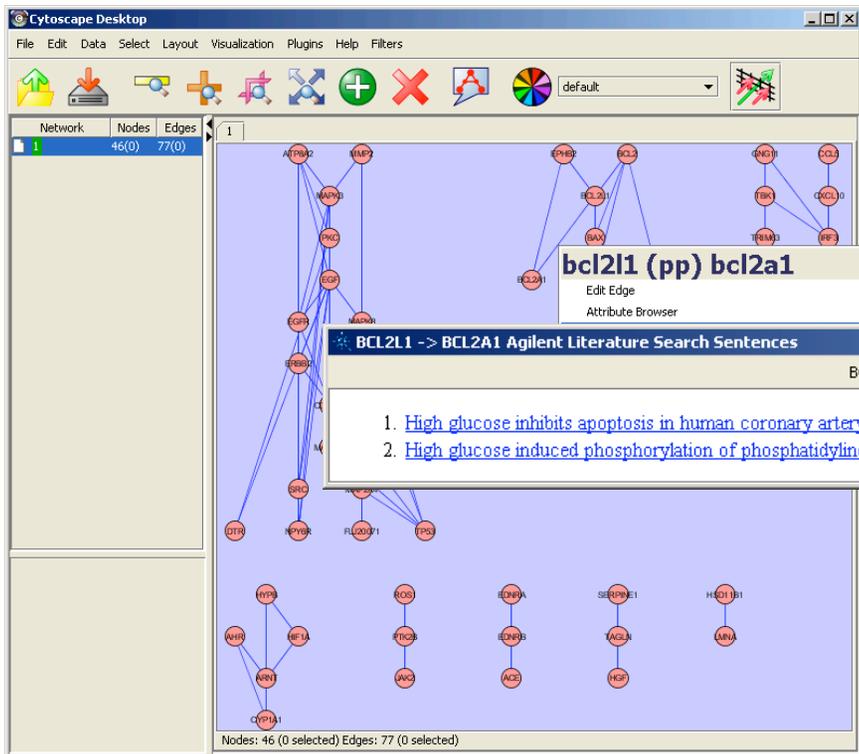
**Results**

- [Association between the eNOS \(Glu298Asp\) and the RAS genes polymorphisms and premature coronary artery disease in a Turkish population \(by Berdeli A, Sekuri C, Sirri Cam F, Ercan E, Sagcan A, Tengiz I, Eser E, Akim M\).](#)  
 BACKGROUND: The renin-angiotensin system (RAS) and endothelial nitric oxide (NO) affect the pathogen...  
 Source:  
 [PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list\_uids=155638  
 75

# Cytoscape Network produced by Literature Search.

Abstract from the scientific literature

Sentences for an edge



Search results window showing a URL: [www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&opt=Abstract&list\\_uids=](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&opt=Abstract&list_uids=)

BCL2L1 -> BCL2A1 Agilent Literature Search Sentences

1. [High glucose inhibits apoptosis in human coronary artery smooth muscle cells by increasing bcl-xL and bfl-1/A1.](#)
2. [High glucose induced phosphorylation of phosphatidylinositol 3-kinase \(PI 3-K\) and extracellular signal-regulated kinase \(ERK\)1/2 along with bcl-xL and bfl-1/A1 upregulation.](#)

Abstract from the scientific literature

High glucose inhibits apoptosis in human coronary artery smooth muscle cells by increasing bcl-xL and bfl-1/A1.

Okumura M, Okumura M, Kojima T, Maruyama T, Yasuda K.

Internal Medicine, Gifu University School of Medicine, Gifu 500-8705, Japan.

Cardiovascular disease is a serious complication in diabetic patients. To elucidate the precise mechanisms of atherosclerosis in diabetic patients, the effects of high glucose concentration (25 mM) on apoptosis regulation and bcl-2 family protein expression in human coronary artery smooth muscle cells (CASMC) were examined. Treatment with a high level of glucose (25 mM) caused a significant decrease in apoptosis in CASMC compared with the same cells treated with a physiologically normal glucose concentration (5.5 mM) (23.9 +/- 2.4% vs. 16.5 +/- 1.8%, P < 0.01). With respect to apoptosis regulation, treatment of CASMC with high glucose concentration markedly increased mRNA expressions of bcl-xL and bfl-1/A1 compared with cells treated with normal glucose. High glucose induced phosphorylation of phosphatidylinositol 3-kinase (PI 3-K) and extracellular signal-regulated kinase (ERK)1/2 along with bcl-xL and bfl-1/A1 upregulation. These results suggest that high glucose suppresses apoptosis via upregulation of bcl-xL and bfl-1/A1 levels through PI 3-K and ERK1/2 pathways in CASMC. High glucose-induced increase in the expression of antiapoptotic proteins may be important in the development of atherosclerosis in diabetic patients.

PMID: 12107051 [PubMed - indexed for MEDLINE]

Display: Abstract Show: 20 Sort by: Send to:

[Write to the Help Desk](#)  
[NCBI | NLM | NIH](#)  
[Department of Health & Human Services](#)  
[Privacy Statement](#) | [Freedom of Information Act](#) | [Disclaimer](#)

Mar 22 2005 17:30:14

# Gene Expression/Network Integration

- Identifier (ID) mapping
  - Translation from network IDs to gene expression IDs e.g. Affymetrix probe IDs
  - Also: Unification, link out, query
  - Entrez gene IDs (genes), UniProt (proteins)
- Synergizer
  - [llama.med.harvard.edu/cgi/synergizer/translate](http://llama.med.harvard.edu/cgi/synergizer/translate)
- More ID mapping services available
  - <http://baderlab.org/IdentifierMapping>

# Gene Expression/Network Integration

## THE SYNERGIZER

The Synergizer database is a growing repository of gene and protein identifier synonym relationships. This tool facilitates the conversion of identifiers from one naming scheme (a.k.a "namespace") to another.

load sample inputs

Select species:

Select authority:

Select "FROM" namespace:

Select "TO" namespace:  [ 854192]

File containing IDs to translate:

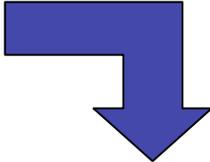
and/or

IDs to translate:

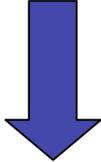
Output as spreadsheet:

Submit

(NB: The strings in [brackets] are representative IDs in the corresponding namespaces.)

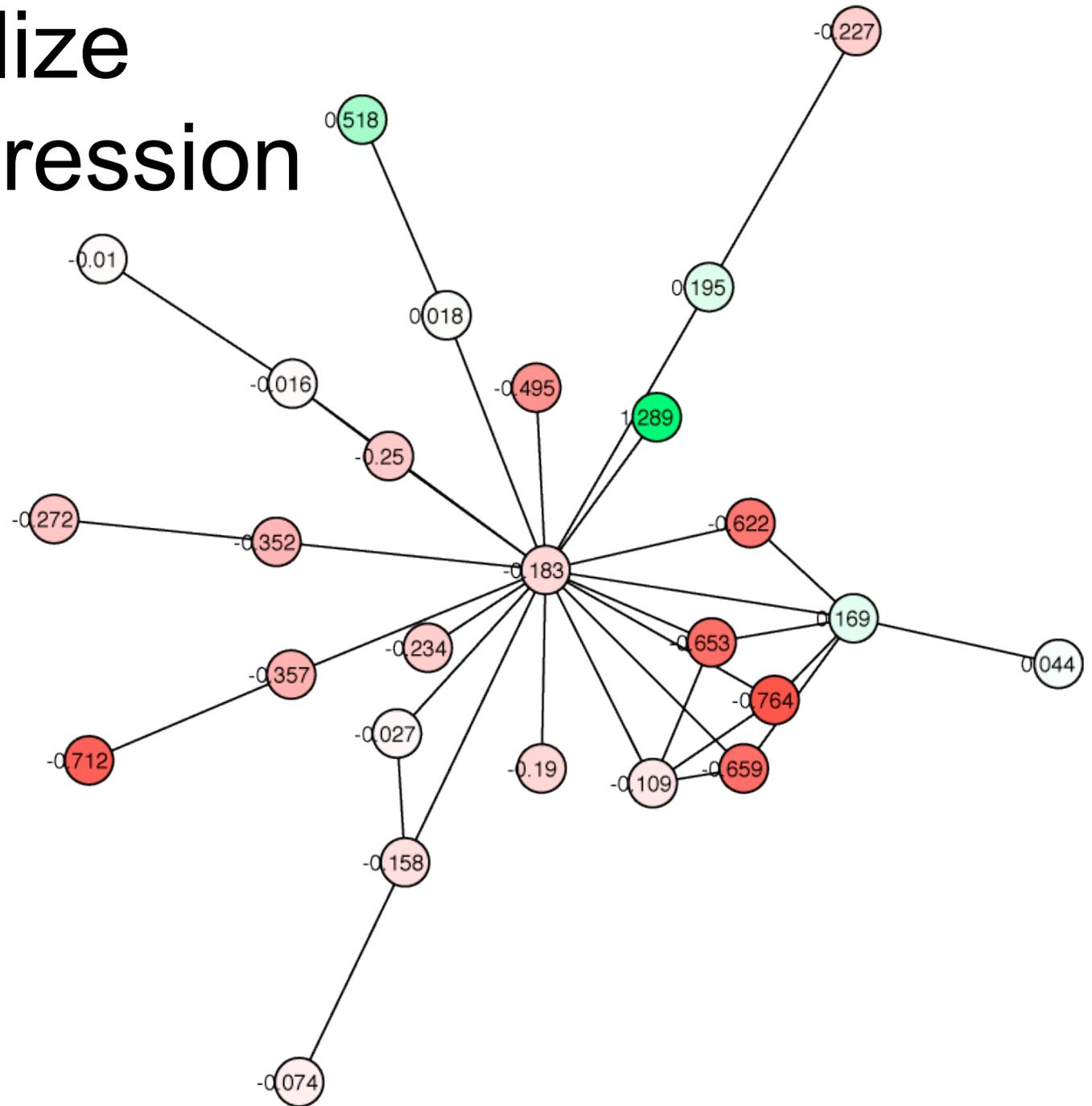


*	entrezgene
YIL062C	854748
YLR370C	851085
YKL013C	853856
YNR035C	855771
YBR234C	852536



1. Load as attributes in Cytoscape
2. Assign expression values to nodes using this attribute set

# Visualize Gene Expression



# Find Active Subnetworks

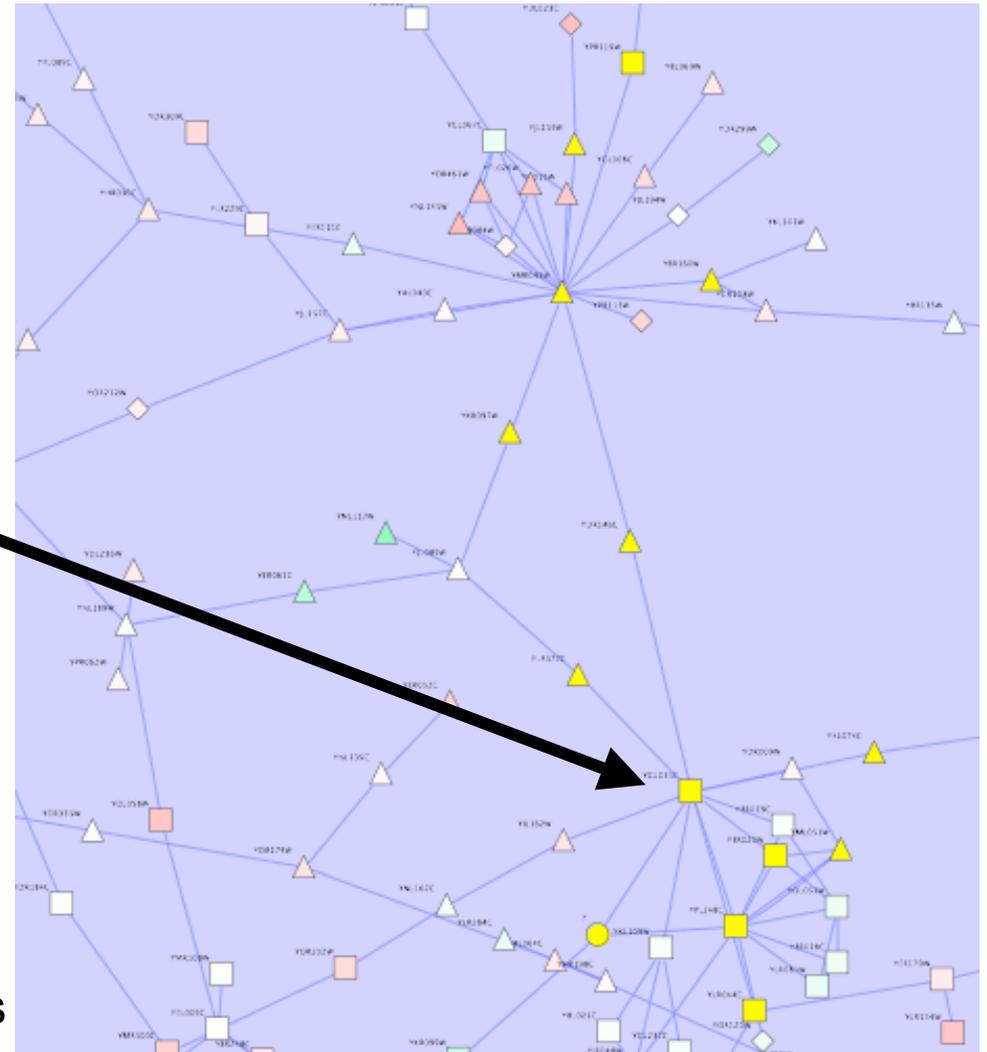
- Active modules
  - Input: network + p-values for gene expression values e.g. from GCRMA
  - Output: significantly differentially expressed subgraphs
- Method
  - Calculate z-score/node,  $Z_A$  score/subgraph, correct vs. random expression data sampling
  - Score over multiple experimental conditions
  - Simulated annealing used to find high scoring networks

# Active Module Results

Network: yeast protein-protein and protein-dna network  
Expression data: 3 gene knock out conditions (enzyme, TF activator, TF repressor)

Network	Size	Score	gal1RGsig	gal4RGsig	gal80Rsig
1	14	3.78			
2	26	3.584			
3	10	2.994			
4	7	2.934			
5	4	2.636			

Save Dismiss



Note: non-deterministic, multiple runs required for confidence of result robustness

Ideker T et al. Science. 2001 May 4;292(5518):929-34.

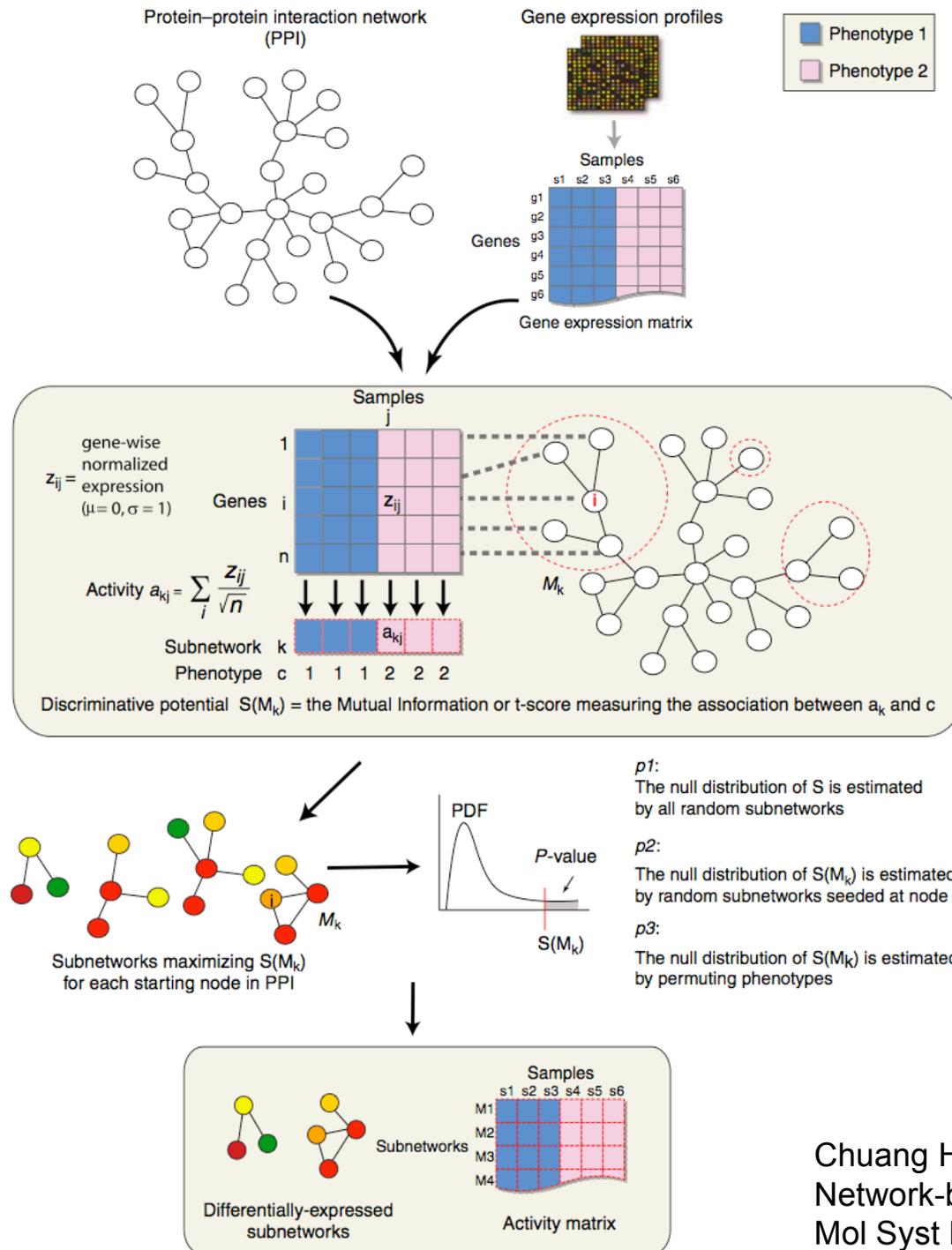
# Network classification of disease

- Traditional: Gene association
- Limitations: Too many genes reduces statistical power
- New: Active cell map based approaches combining network and molecular profiles
- <http://www.connotea.org/user/ACMT>

Chuang HY, Lee E, Liu YT, Lee D, Ideker T  
Network-based classification of breast cancer metastasis  
Mol Syst Biol. 2007;3:140. Epub 2007 Oct 16

Liu M, Liberzon A, Kong SW, Lai WR, Park PJ, Kohane IS, Kasif S  
Network-based analysis of affected biological processes in type 2 diabetes models  
PLoS Genet. 2007 Jun;3(6):e96

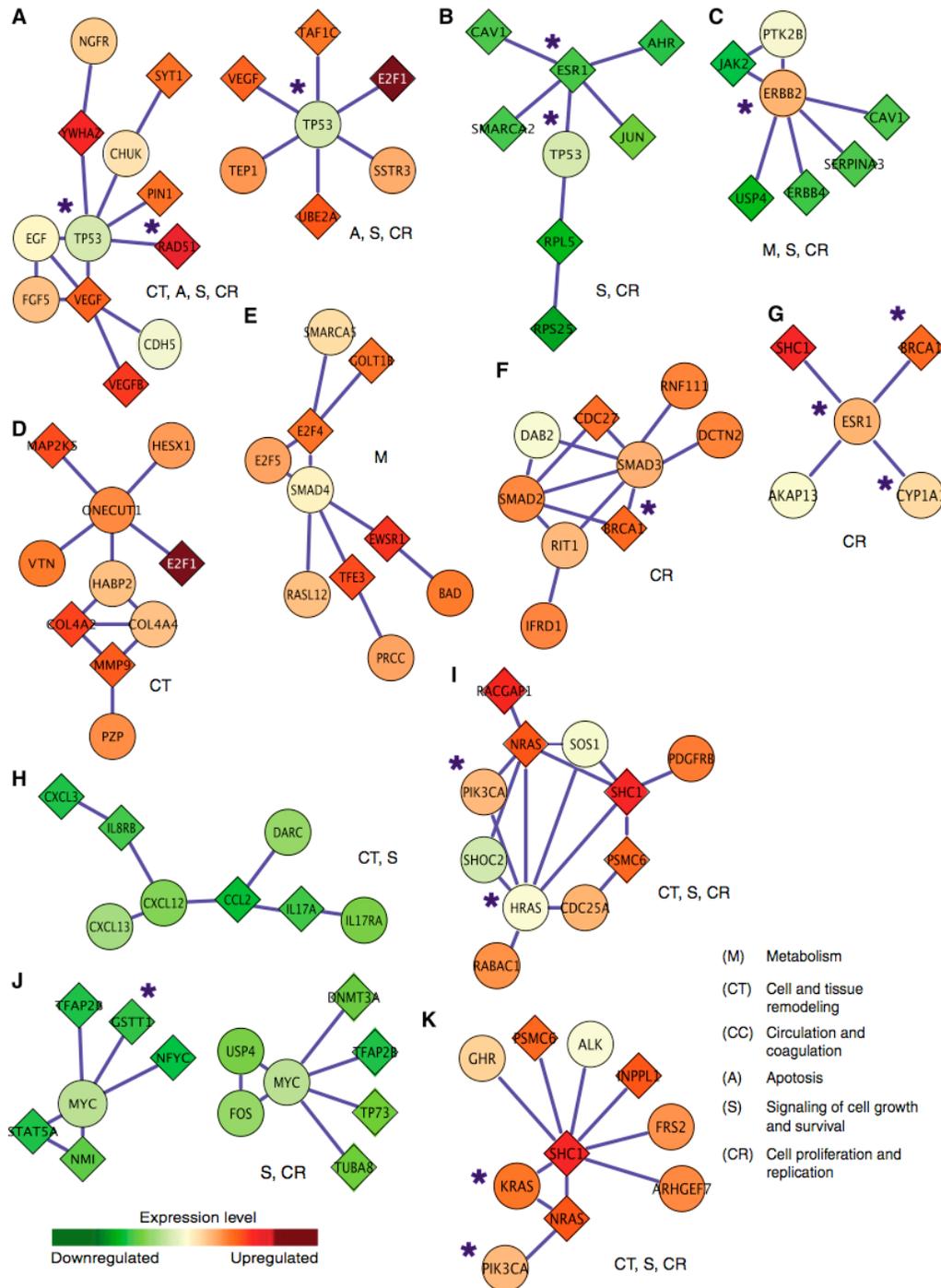
Efroni S, Schaefer CF, Buetow KH  
Identification of key processes underlying cancer phenotypes using biologic pathway analysis  
PLoS ONE. 2007 May 9;2(5):e425



## Network-based breast cancer classification

- 57k intx from Y2H, orthology, co-citation, HPRD, BIND, Reactome
- 2 breast cancer cohorts, different expression platforms
- Metastasis: 78 vs. 217 (van de Vijver) & 106 vs. 180 (Wang)

Chuang HY, Lee E, Liu YT, Lee D, Ideker T  
 Network-based classification of breast cancer metastasis  
 Mol Syst Biol. 2007;3:140. Epub 2007 Oct 16



- Similar network markers across 2 data sets (better than original overlap)
- Increased classification accuracy
- Better coverage of known cancer risk genes (\*)

# Acknowledgements

## Pathway Commons

**Chris Sander**

Ethan Cerami

Ben Gross

Emek Demir

Robert Hoffmann

## Cytoscape

**Trey Ideker (UCSD)**

Ryan Kelley, Kei Ono, Mike Smoot, Peng

Liang Wang (Nerius Landys, Chris

Workman, Mark Anderson, Nada Amin,

Owen Ozier, Jonathan Wang)

## Lee Hood (ISB)

Sarah Killcoyne (Iliana Avila-Campillo,

Rowan Christmas, Andrew Markiel, Larissa

Kamenkovich, Paul Shannon)

## Benno Schwikowski (Pasteur)

Melissa Cline, Tero Aittokallio

## Chris Sander (MSKCC)

Ethan Cerami, Ben Gross (Robert Sheridan)

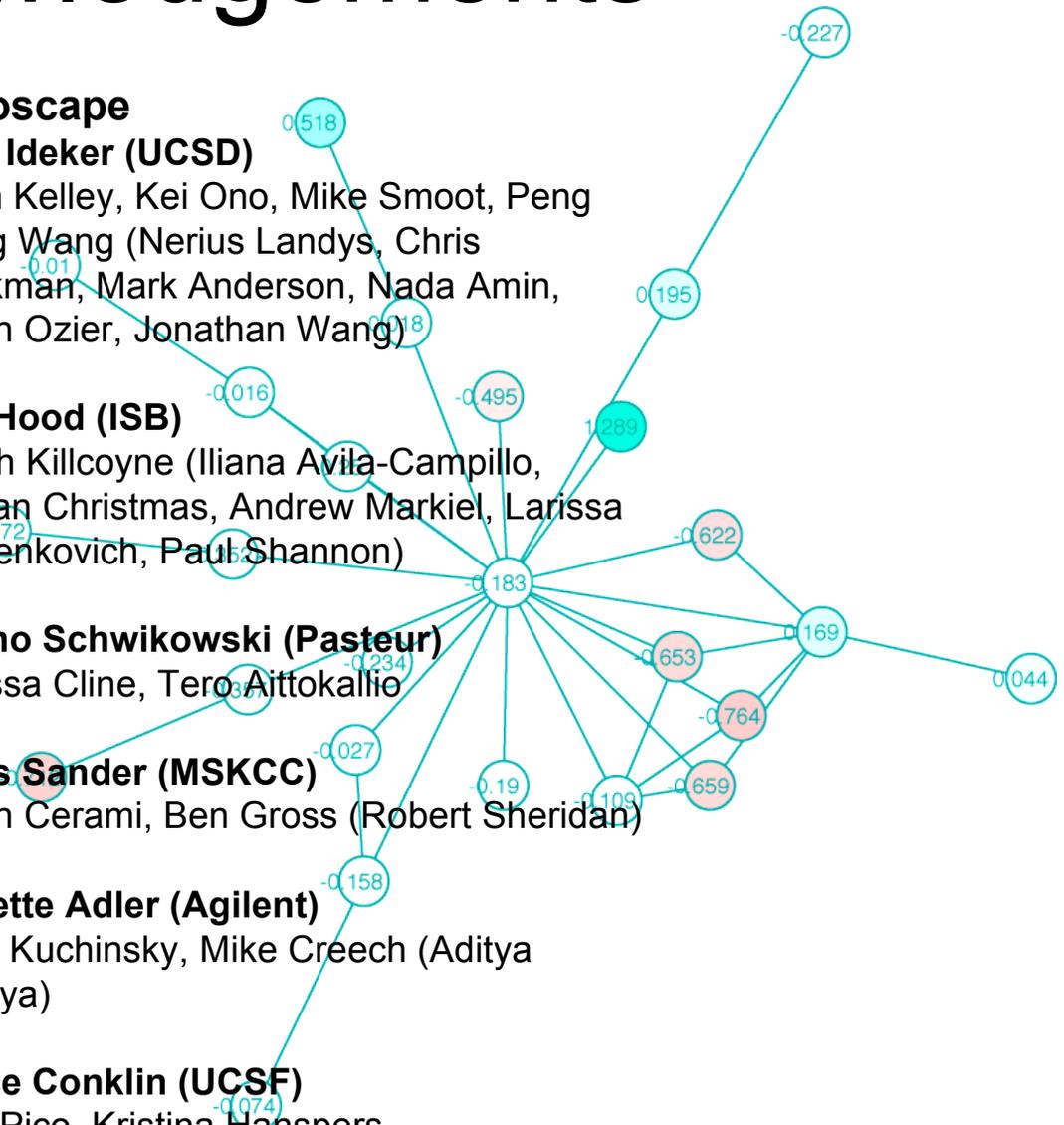
## Annette Adler (Agilent)

Allan Kuchinsky, Mike Creech (Aditya

Vailaya)

## Bruce Conklin (UCSF)

Alex Pico, Kristina Hanspers



## Funding

CIHR

Genome Canada

Canada Foundation

for Innovation

<http://baderlab.org>